Allostery and Intrinsic Disorder in Ensembles and Single Molecules

by
Hesam N. Motlagh

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
September 2015

Abstract

Life requires regulated responses to signals. Without tightly-controlled regulation, multi-cellularity and the development of life are seemingly intractable. Allostery, i.e. the binding of a molecule distal to the site it effects, is a ubiquitous mechanism employed by nature to tackle such challenges. For over 50 years the reigning paradigm has been a structural view of allostery, i.e. one that focuses on a mechanical site-to-site perturbation in a biological macromolecule that results in the allosteric effect. However, what has come to light the past two decades is that nature utilizes the full repertoire of dynamics and disorder in propagating the allosteric signal. How do we reconcile these disparate phenomena? This thesis draws attention to the statistical nature of biological macromolecules, and utilizes such a line of inquiry to glean insight into the ground rules of allostery. We first discuss a theoretical framework that can generally treat allosteric phenomena, which then motivates experiments on intrinsically disordered proteins. Due to the ensemble nature of bulk experiments, we built and calibrated a single-molecule force spectroscopy instrument to directly probe the statistical nature of such biological processes at the single-molecule level. We found that by perturbing solution conditions, we not only gain insight into structured proteins, but also force disordered proteins to fold to their biologically active state(s). The data are rich in information and suggest that the active state of disordered proteins are multi-state lending credence to their statistical nature. Taken together, these results illuminate the ensemble nature of allostery and provide a vehicle for reconciling allosteric phenomena in disparate systems.

Advisor: Vincent J. Hilser, PhD

Reader: Doug Barrick, PhD

Committee: Bertrand Garcia-Moreno E, PhD

James Berger, PhD

Elijah Roberts, PhD

Preface

There is no way to do justice to immeasurable assistance, mentorship, encouragement, and love I have received throughout my life to make this dissertation possible. The following is an attempt to thank the many people who have helped me along this journey

I would like to begin by thanking the faculty and staff of the T. C. Jenkins Department of Biophysics and the Department of Biology for their assistance in my graduate education. In particular, I thank Dr. Bertrand Garcia-Moreno as he has served as one of my greatest intellectual influences throughout my time here at Hopkins. I would also like to thank Ranice Crosby, Kathy Kolish, and Nicole Goode for being the "mothers" of our program – they are the reason why the biophysics community is so strong. My thesis committee members, Dr. Doug Barrick, Dr. James Berger, and Dr. Elijah Roberts also deserve acknowledgment and thanks for their mentorship and guidance. And of course, Dr. Greg Bowman, for not only serving as a previous member and current alternate for my committee, but for teaching me how to critically read a scientific paper (whether I liked it or not) in his proteins and nucleic acids course

My dissertation would not have been possible without those in my personal life who have supported me for so many years. The biophysics and graduate student community is the reason why I did not drop out of graduate school a few years ago. The Hopkins Brewing and Distilling Team will always have a special place in my heart – Dr. Robert Trachmann, Pat Byrne, Andy Goodrich, John Froehlig, Mr. White, Kiara Eldred, William Keenan, Anthony Milin, and Stella Eldred. I must thank Scott Nichols and Alison Ringel as well – they are the reason I have become my own worst critic, a necessity as a scientist and critical thinker. I also must thank the Johns Hopkins University Ice Hockey team for their support and stress-relief – they have even placed their trust in me by letting me captain them into battle the past three years.

Dr. Vince Hilser has been an excellent graduate mentor. The research environment in the Hilser lab provided me with time, space, financial support, and intellectual drive to pursue my own independent ideas even if Vince did not agree with them. Vince gives his students a long leash, sometimes long enough that you can hang yourself with it, however, he unconditionally supports his students and treats them as an equal. I am now beginning to realize how he has taught me to think *differently* at a level that has made me a better scientist, and more importantly, a better person. I would also like to acknowledge past and current members of the Hilser lab for creating a great learning environment: Dr. Austin Elam, Dr. Jing Li, Dr. Harry Saavarda, Jordan White, Alex Chin, Andrew Martens, James Rives, Joe Rehfus, Emily Grasso, and Jeremy Anderson.

Lastly, I would like to acknowledge my family who has helped make this thesis possible with their love and support. During my graduate career I received constant support from my parents, Mahnaz and Hossein, and my sister, Mahsaw. I dedicate my thesis to these three.

Table of Contents

Abstract ii

Preface iii

Table of Contents v

List of Tables ix

List of Figures x

Chapter 1 – Introduction 1

1.1 The ensemble nature of allostery 1

1.2 The ensemble allosteric model 2

1.2 Single-molecule force spectroscopy illuminates the ensemble nature of proteins 4

1.4 The interplay between intrinsic disorder and allostery 5

1.5 Overview of the thesis 6

Chapter 2 – Statistical thermodynamics of allostery and intrinsic disorder 11

2.1 Abstract 11

2.2 Introduction 12

2.3 Materials and Methods 14

Theoretical derivation of a statistical thermodynamic framework for allostery 14

Exhaustive search of thermodynamic parameter space 18

2.4 Results and Discussion 19

Results of the Model 19

Thermodynamic basis for agonism/antagonism switching 21

Implications of the model 23

Relationship to classical models of allostery 24

2.5 Conclusions 25

Chapter 3 – Single-molecule force spectroscopy instrumentation and calibration 37

3.1 Abstract 37

3.2 Introduction 38

3.3 Materials and Methods 39

Optical MiniTweezers instrumentation 39

Single-molecule DNA handle generation 40

Direct measurement of entropy from thermal noise 41

3.4 Results and Discussion 42

Force Calibration 42

Distance Calibration 45

Single-molecule force-spectroscopy on 3kbp DNA handles 47

Experimental “violations” of the 2nd law of thermodynamics 48

3.5 Conclusions 50

Chapter 4– Two-dimensional single-molecule force spectroscopy allows structural modeling of protein folding 60

4.1 Abstract 61

4.2 Introduction 62

4.3 Materials and Methods 64

Pulling Buffers and Materials 64

Protein purification 64

Synthesis of Oligonucleotide crosslinked to CoA 65

Labeling of protein for optical tweezer experiments 65

Synthesis of DNA handles 67

Synthesis of beads for protein ligation 67

Optical tweezers measurements and analysis 68

Derivation of transfer free energy for intermediate states 68

Steered molecular dynamics simulations 71

Structure based calculations for the intermediate state 72

4.4 Results and Discussion 73

Osmolytes minimally affect the mechanical unfolding pathway 73

Osmolytes accelerate early events in folding 76

Unique access to the intermediate state characterization 79

4.5 Conclusions 83

Chapter 5 – Single-molecule force spectroscopy on disordered proteins in the presence of osmolytes 95

5.1 Abstract 96

5.2 Introduction 97

5.3 Materials and Methods 99

TMAO induced protein folding transitions 99

Assessment of side chain and backbone contributions to the m-value for TMAO-induced folding 99

Protein purification for single-molecule experiments 100

5.4 Results and Discussion 101

Intrinsically disordered proteins are thermodynamically similar to globular proteins 101

Intrinsically disordered proteins are amenable to single-molecule force spectroscopy 103

Glucocorticoid Receptor's N-terminal domain is either misfolded or multi-state 106

5.5 Conclusions 109

References: 123

Vita 132

List of Tables

Table 2-1 - Agonism/Antagonism switching behavior is encoded in the thermodynamic architecture and not qualitatively dependent on ligand binding free energy. 35

Table 2-2. Point density of Regions 1 and 2 show Region 2 is more often populated. 36

Table 3-1. Calibration constants for optical MiniTweezers instruments demonstrate proper calibration. 59

Table 4-1 – Unfolding kinetics upon addition of osmolyte 94

Table 4-2 – Folding kinetics upon addition of osmolyte 94

Table 5-1. Assessment of m-value contributions from protein backbone (BB) and side chains (SC) 121

Table 5-2. Statistics on single-molecules of A-NTD and their propensity for structural transitions under different conditions. 122

List of Figures

Figure 1-1. The dynamic continuum of allostery. 8

Figure 1-2. The ensemble allosteric model accounts for the intrinsic energetics of allosteric proteins. 9

Figure 2-1. Three-domain allosteric protein. 27

Figure 2-2. Specific thermodynamic architectures can produce agonism/antagonism switching. 29

Figure 2-3. Agonism/Antagonism switching is highly degenerate for stabilities of domains. 31

Figure 2-4. Energetic rules for agonism-antagonism switching. 32

Figure 2-5. Switching-competence is maximized when regulatory domains are predominantly populating low affinity or ID states. 34

Figure 3-1. General overview of the optical MiniTweezers instrument. 52

Figure 3-2. Laser optical path for the optical MiniTweezers. 53

Figure 3-3. Force detection depends on the angle of incident light. 54

Figure 3-4. Force calibration representative data. 55

Figure 3-5. Light lever calibration representative data. 56

Figure 3-6. Single-molecule force spectroscopy on DNA handles reveals proper calibration of instrumentation. 57

Figure 3-7. Experimental “violations” of the second law of thermodynamics. 58

Figure 4-1. T4 Lysozyme as a model system and single molecule folding experimental setup 85

Figure 4-2. The mechanical unfolding pathway is unaffected by osmolyte 86

Figure 4-3. Osmolytes affect protein folding kinetics 87

Figure 4-4. Osmolytes specifically affect the first step in folding. 88

Figure 4-5. Osmolytes stabilize formation of the intermediate. 89

Figure 4-6. Backbone exposed surface area upon mechanical unfolding. 90

Figure 4-7. Side-chain exposed surface area upon mechanical unfolding. 91

Figure 4-8 – Structural model of the intermediate and its predictive power..... 92

Figure 5-1. Glucocorticoid Receptor domain architecture and naturally occurring translational isoforms. 110

Figure 5-2. Glucocorticoid receptor N-Terminal Domain isoform stability contributes to *in vivo* activity and refolds to a globular conformation. 111

Figure 5-3. The refolded state(s) of disordered proteins are thermodynamically similar to globular proteins. 113

Figure 5-4. Single-molecule force spectroscopy experimental setup for intrinsically disordered proteins..... 114

Figure 5-5. Intrinsically disordered proteins do not exhibit tertiary structural transitions under mechanical force 115

Figure 5-6. Certain isoforms of an intrinsically disordered protein exhibit mechanical unfolding transitions upon addition of osmolyte. 116

Figure 5-7. The TMAO refolded state(s) of A-NTD are heterogeneous when mechanically unfolded. 117

Figure 5-8. GR's NTD is misfolded upon addition of osmolyte..... 118

Figure 5-9. States of A-NTD is too heterogeneous to assign even in the presence of proline isomerase. 120

Chapter 1 – Introduction

1.1 The ensemble nature of allostery

One of the greatest challenges biological systems encounter is changes in their environment [1]. As a result, the development and propagation of biological life requires regulated responses to environmental and internal cues. Without controlled and fine-tuned regulation in biology, critical processes such as metabolism, multicellularity, and development are seemingly intractable [2]. A ubiquitous mechanism that nature utilizes to tackle these processes is allostery [3-9]. Allostery is the regulation of a molecule by a ligand binding distal from the site it affects, deriving from the Greek *allos* meaning "other" and *stereos* meaning "structure". Allostery has even been referred to as the "second secret of life" [7]. Furthermore, dysregulation of allostery can lead to diseases and cancers, lending credence to its critical importance in biological function and the necessity to understand it at a fundamental level [9, 10]. Given its ubiquity and significance, it is of no surprise that many efforts have focused on studying this process for over 50 years [11-13].

Initial efforts to understand the mechanism of allostery focused on the structural basis of effects in hemoglobin [14-16]. Through the lens of history, it is clear that the first models of allostery were proposed in a time in which the first high-resolution structures were being solved which guided the thinking in the field [17]. Although the structural view of allostery has been used to explain mechanism in multiple systems and may even be the dominant form of signal propagation [18, 19], what has become clear over the past few decades is that this view does not explain allostery in all systems [6]. Specifically, the classic structure-based allosteric mechanisms [16, 20-22] must be cast within the context of new discoveries, revealing that allostery can be manifested by changes in protein dynamics and conformational disorder or unfolding [23-26]. Indeed, recent studies have shed light on how nature utilizes the full repertoire

of conformational heterogeneity [27] (*e.g.* rigid body movement [19, 28, 29], folded yet dynamic structure [30], and intrinsic disorder [31]) to facilitate allostery. We have dubbed this spectrum of phenomena the "dynamic continuum" of allostery (Fig. 1-1) [32].

How do we quantitatively reconcile allosteric phenomena? The two classic models of allostery that have attempted to address this question are the, "symmetric," Monod-Wyman-Changeux (MWC) model [33] and the, "sequential," Koshland-Nemethy-Filmer (KNF) model [34]. Both models have demonstrated great utility with the MWC model being the dominant model over the course of time [3]. However, it is now well-appreciated that both of these models are phenomenological and do not provide mechanistic insight into "how" allostery is encoded structurally, kinetically, or thermodynamically [8]. Furthermore, these models fail to capture all processes on the dynamic continuum of allostery (Fig. 1-1). What is clear is that the structural view of allostery needs to shift to a line of questioning that focuses on the statistical nature of the allosteric process and the signal propagation across molecules within the context of heterogeneity to fully capture the entire dynamic continuum. This is what we refer to as the ensemble view of allostery [32].

1.2 The ensemble allosteric model

In an effort to generalize the notion of allostery, Hilser and Thompson have proposed an Ensemble Allosteric Model (EAM) that is rooted in the intrinsic energetics of allosteric proteins [35]. The model quantitatively treats allosteric protein with three well-established assumptions: (1) that the protein exists in a conformational ensemble, the hallmark of an allosteric protein [8, 32, 36, 37]; (2) the binding of ligand occurs to modular and individual domains [38]; (3) and that the domains are energetically coupled to one another, i.e. the essence of allosteric effect [39] (Fig. 1-2A). The model articulates the intrinsic free energies for the protein based on these assumptions

which yields a statistical thermodynamic framework that describes allosteric phenomenon. Importantly, this model is distinct from the MWC and KNF models which can provide insight into new allosteric phenomena [8].

Since the EAM is articulated in general energetic terms, the model can be agnostic to structure. For instance, in the simplest case, if there are two conformations for each domain, they can be treated as two ostensibly structured folded conformations or an unfolded to folded transition (Fig. 1-2B). The important aspect is that there is a binding affinity difference between these states which drives the allosteric response. Preferential binding affinity to one conformation or the other will bias the distribution of states in the ensemble [40]. What is clear is that this model leads to conclusions that neither the MWC nor KNF models do. Specifically, the EAM leads to the conclusion that allosteric coupling between domains is optimized when they are intrinsically disordered (ID) in the absence of ligand, i.e. the regulatory domains are unfolded a vast majority of the time when we consider a folded-unfolded transition [41]. This is counterintuitive to the structural view of allostery. If the allosteric signal is propagated by a well-defined structural pathway, how is possible that intrinsic disorder (i.e. a lack of structure) optimizes the ability to couple domains? Results like these demonstrate the utility in being agnostic to structure and may explain the role of intrinsic disorder in propagating allosteric signals, as is discussed in further detail in the following sections.

Perhaps the most important insight provided by the EAM is that it is transferable across the dynamic continuum of allostery. As such, this model can provide a general framework for allostery in all systems and can even reconcile certain paradoxical allosteric observations. For instance, the same allosteric effector can act as an positive effector (agonist) under one set of conditions and an negative effector (antagonist) under another set of conditions for ordered [42] and disordered proteins [43]. In Chapter 2, the EAM is used to address this paradoxical observation. The ensemble nature of allostery illuminated from such a framework suggests that

intrinsic disorder can even be used to evolve systems that can be *functionally pluripotent* – i.e. can up- or down-regulate in response to external stimuli without being a fixed molecular switch. These results motivated the experimental portion of this thesis, i.e. to explore the ensemble nature of allostery in disordered proteins.

1.2 Single-molecule force spectroscopy illuminates the ensemble nature of proteins

The ensemble nature of allostery discussed above suggests that mechanistic insight at the experimental level will require deconvolution of most, if not all, relevant states populated under different conditions. An issue with bulk-experimental methods is that they represent an ensemble average, and therefore do not provide direct access to lowly populated states [44]. Furthermore, experimental analysis of IDPs can be limited by solubility which can lead to off-pathway aggregation [45]. Single-molecule methods mitigate both of these issues and provide an avenue to experimentally dissect the ensemble nature of allostery.

Single-molecule force spectroscopy has emerged as a powerful tool to study protein-folding [46] and determining functional sub-domains [47]. The ability to resolve single-molecule trajectories has revealed a wealth of information on intermediates that are biologically meaningful [47]. Towards the goal of performing such experiments to dissect the ensemble nature of allostery in disordered proteins, we opted to build a laser optical tweezers instrument. In Chapter 3, a brief overview of the instrumentation and calibration for an optical MiniTweezers is discussed. Since such instruments are not commercially available they must be built in-house and maintained daily.

In Chapter 4, the instrumentation was tested on a well-characterized system: T4 Lysozyme. We obtained results that were in excellent agreement with previously published data

[48]. Furthermore, it was clear from previous results that single-molecule experiments on our IDP model system would result in multi-state behavior. How do we deconvolve intermediate species observed in single-molecule traces without mutagenesis? We tested an approach by using naturally occurring solutes called osmolytes, which reports on the relative surface area of states [49-51]. We found that with T4 lysozyme, we are able to perform structure based energy calculations to deconvolve at the amino acid level what an intermediate may look like by modulating the energy landscape with known concentrations of osmolytes. We hoped to use this strategy when we performed experiments with IDPs, specifically that the heterogeneous amino acid composition would inform which segments of the molecule are undergoing structural transitions. Regardless of the lack structural information on IDPs, these results suggest that we should be able to rationally predict the sensitivities of specific subdomains based on amino acid composition alone. This would aid in state assignment and demarcation of domains observed as intermediates. Generally, this strategy may provide a means of deconvoluting intermediates and equilibrium states in all single molecule experiments.

1.4 The interplay between intrinsic disorder and allostery

Recently there has been a realization that ID regions (IDRs) and proteins (IDPs) are hyper-abundant in transcription factors and cell signaling pathways, specifically in higher organisms [52]. Given that these processes specifically need tightly-regulated response from external signals, suggests that IDRs and IDPs encode the ability for allostery. Until recently, experimental evidence for this hypothesis has been sparse [43, 53, 54]. Perhaps most illuminating has been a recent single-molecule study that has revealed how these IDPs endow the ability for even cooperative “switching” behavior [43] as predicted by the EAM [8, 32, 55].

Our lab has been responsible for thermodynamically dissecting a disordered protein, the N-terminal domain (NTD) of *H. sapiens* glucocorticoid receptor (GR) [54]. Previous members of the lab have demonstrated that the NTD consists of allosterically coupled sub-domains that modulate the net instability of the NTD, and hence, contribute to *in vivo* activity. In Chapter 5, we discuss how osmolyte induced folding experiments and structure-based energy calculations reveal that the refolded state(s) of the NTD and its various isoforms are thermodynamically indistinguishable from its globular protein counter parts. This led to the hypothesis that the osmolyte induced conformation(s) of GR may be mechanically unfolded using single-molecule force spectroscopy similar to globular proteins.

We hypothesized that these experiments would reveal a wealth of information on the coupling between sub-domains, their boundaries in amino acid space, and the relative kinetics or thermodynamics of all species populated in the ensemble. We find that IDPs are amenable to single-molecule force spectroscopy, and what may appear to be a two-state transition in bulk, is in fact multi-state at the single molecule level. The data reveal that the NTD is multi-domain and too complicated to be state assigned reliably on its own. Further experimental avenues to pursue are mentioned and the current results are cast within the context of the IDP field.

1.5 Overview of the thesis

The work in this thesis illuminates the ensemble nature of allostery. Whether it is in the theoretical framework of the EAM in Chapter 2, or the single-molecule force spectroscopy experiments in Chapters 3-5, what is clear is that these results shed light on the statistical nature of the underlying biological process. These results suggest a new line of questioning – i.e. one that changes the focus from structural pathways through proteins, to one that focuses on the

coupled segments of the molecule and their relative stabilities. It is this type of framework that generates the delicate balance of states populated by biological macromolecules. This of course does not represent a new view, in fact the authors of the MWC [33] and KNF [34] models were very aware of the ensemble nature of molecules. Furthermore, the structural view of allostery has provided immense intellectual strides forward that have been used to glean insight into therapeutic design of drugs [56]. However, the static view of proteins must be reassessed given the wealth of experimental data that cannot be explained by structure alone. What is clear, is that future research endeavors will have to focus on the ensemble nature of allostery to glean mechanistic insight into the underlying biological process, and inevitably, design of allosteric therapeutics.

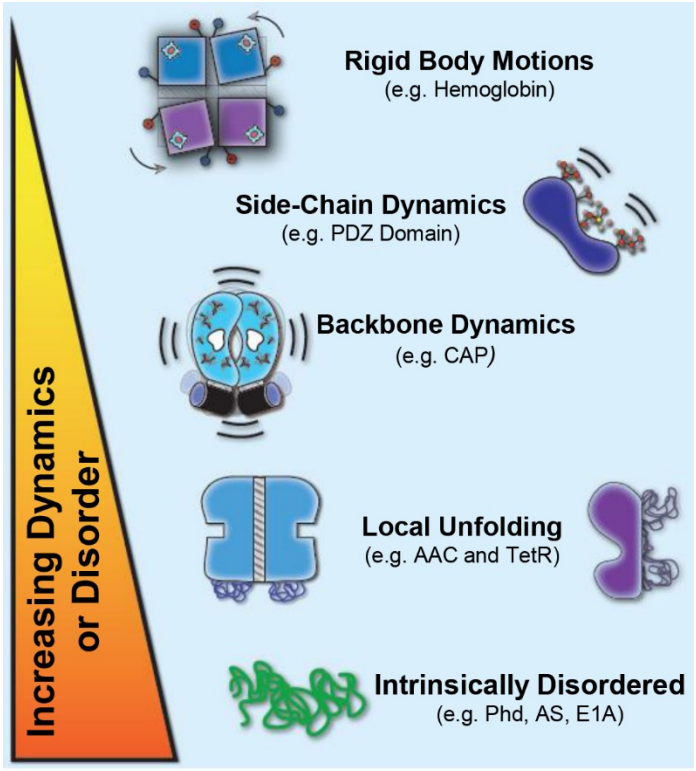

Figure 1-1. The dynamic continuum of allostery. Shown is a schematic representation of allosteric phenomena with increasing dynamics or disorder near the bottom. Specific examples shown are rigid body motions in hemoglobin, side-chain dynamics in a PDZ domain [30], backbone dynamics in catabolite activator protein (CAP) [57-59], local unfolding in the enzyme AAC [60] and the transcription factor TetR [61], and intrinsically disordered allostery as mentioned in the text. Figure was adopted from [32].

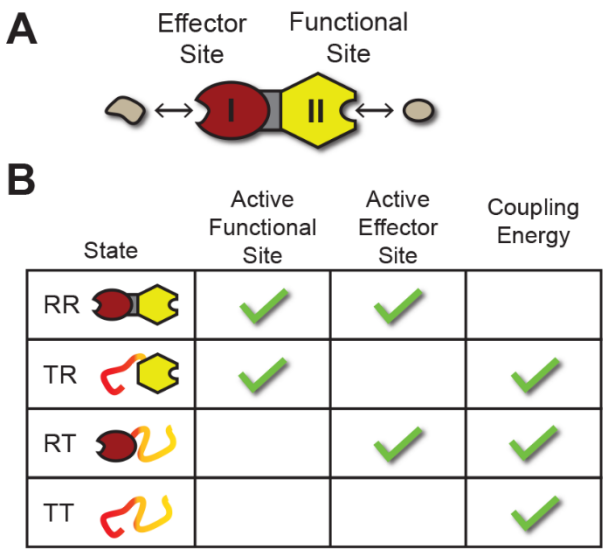

Figure 1-2. The ensemble allosteric model accounts for the intrinsic energetics of allosteric proteins. A. Schematic representation of a two-domain allosteric protein with an effector site (domain I, red) and a functional site (domain II, yellow). Both sites can bind their respective ligands. The two domains are coupled to one another as depicted by the gray shaded bar. **B.** A schematic of the ensemble of states accessible to the two-domain allosteric protein. Each domain can be in a high-affinity relaxed (R) or a low-affinity tensed (T) state resulting in the four states depicted in the first column: RR, TR, RT, and TT. The T state is represented as unstructured (i.e. locally unfolded), however, the model can treat two ostensibly folded conformations as long as there is a difference in binding affinity. Columns 2-3 show whether the functional or effector site is active (i.e. in the R-conformation). Column 4 shows whether the coupling energy is broken in the state which can be appreciated by the lack of the gray shaded bar in the RR-state. In short, each state has its own unique properties and the relative probabilities of each state are determined by the intrinsic energetics. Figure adopted from [41].

Chapter 2:

Statistical thermodynamics of allostery and intrinsic disorder

by

Hesam N. Motlagh[a] and Vincent J. Hilser[*a,b]

[a] T.C. Jenkins Department of Biophysics and [b] Department of Biology,

The Johns Hopkins University,

Baltimore, MD 21218

*To whom correspondence should be addressed: hilser@jhu.edu

This work published as:

Motlagh HN and Hilser VJ. (2012) Agonism/antagonism switching in allosteric ensembles. Proceedings of the National Academy of Science USA, **109** (11) 4134-39.

Chapter 2 – Statistical thermodynamics of allostery and intrinsic disorder

2.1 Abstract

Ligands for several transcription factors can act as positive effectors (agonists) under some conditions and negative effectors (antagonists) under others. The structural and molecular bases of such effects are unknown. Previously, we demonstrated how the folding of intrinsically disordered protein sequences, in particular, and population shifts, in general, could be used to mediate allosteric coupling between different functional domains, a model that has subsequently been validated in several systems. Here it is shown that population redistribution within allosteric systems can be used as a mechanism to tune protein ensembles such that a given ligand can act as both an agonist and an antagonist. Importantly, this mechanism can be robustly encoded in the ensemble, and does not require that the interactions between the ligand and the protein differ when it is acting either as an agonist or an antagonist. Instead, the effect is due to the relative probabilities of states prior to the addition of the ligand. The ensemble view of allostery that is illuminated by these studies suggests that rather than being seen as switches with fixed responses to allosteric activation, ensembles can evolve to be "functionally pluripotent", with the capacity to up or down regulate activity in response to a stimulus. This result not only helps to explain the prevalence of intrinsic disorder in transcription factors and other cell signaling proteins, it provides important insights about the energetic ground rules governing site-to-site communication in all allosteric systems.

2.2 Introduction

The paradigm that proteins acquire their function by adopting a well-defined structure has been challenged by the observation that many proteins are either intrinsically disordered (ID) under native conditions or that there is significant conformational heterogeneity within an otherwise folded protein [52, 62-65]. Of particular significance is the growing body of literature that shows allostery can be accompanied by significant changes in conformational fluctuations [57, 60, 61, 66-71], and that ID sequence stretches are found in hyper-abundance in transcription factors (TFs)[52]. Previously, we demonstrated that folding of intrinsic disorder can be used as a mechanism to couple binding between different functional domains [41]. Our results revealed that the precise distribution of states within the native state ensemble of allosteric proteins was responsible for the magnitude of the allosteric response. In addition, the response could be modulated by tuning the ensemble through mutation or through the binding of ligands, protons or cofactors. This model has subsequently provided insight into allosteric mechanism in both real and designed systems [60, 72, 73]. Here we address the question of whether such tuning of the ensemble can involve transforming an antagonist into an agonist (or vice versa). Answering this question will not only provide insight into allosteric mechanism [74], it could also have a broad impact on strategies currently being implemented in the development of anticancer or other agents that act by allosterically binding to their targets [75].

Most TFs have a modular structure, which includes a regulatory domain that allosterically activates or represses gene expression in response to ligand binding [76]. Historically, ligands that bind to TFs have been classified by their mutually exclusive modes of action: “agonists” if they enhance gene expression or “antagonists” if they repress gene expression [75]. Of interest are the cases that shatter this notion by acting as agonists under one set of conditions, and antagonists under others [77-79]. In one well-documented case, Tamoxifen was developed as an antagonist of estrogen receptor (ER) and acted as such against cancer in

breast tissue. However, Tamoxifen had unanticipated agonistic effects in bone and uterus tissue, resulting in patients having up to a seven-fold increase in risk for endometrial cancer [42]. This phenomenon has qualitatively been explained by differing sets of binding partners that interact with ER in different tissues, yet a quantitative model describing the underlying basis of agonism/antagonism switching is lacking.

The observation that a ligand can act as an agonist and an antagonist for the same protein depending on its context is difficult to explain using the classic Monod, Wyman, Changeux [MWC] [13] or Koshland, Nemethy, and Filmer [KNF] [80] models of allostery because neither model addresses "how" binding at one site affects other sites [6] . Here we show that a general ensemble description of proteins [41] can be used to investigate agonism/antagonism switching in allosteric systems. Within the context of this model, the allosteric response of an ensemble is determined by the local equilibrium constants between high and low affinity states for different ligands, as well as the energetic coupling between the different regions. It is demonstrated, through an unbiased search of thermodynamic parameter space, that the ability to switch between agonistic and antagonistic responses for a given ligand; 1) can be robustly encoded, 2) has simple selection rules based on the interactions between domains, and 3) is maximized when at least one regulatory domain is either disordered or significantly populating a low affinity state. More generally, the results reveal that the ensemble allosteric model (EAM) described here and elsewhere [41] provides a vehicle for interpreting experiments and developing an understanding of the structural basis for allostery in terms of thermodynamic ground rules that govern "how" allostery works.

2.3 Materials and Methods

Theoretical derivation of a statistical thermodynamic framework for allostery

It is commonly recognized that multi-domain regulatory proteins, such as TFs, have a modular structure and often segregate binding sites for each ligand into discrete structural domains [38, 76, 81]. As such, an allosteric protein can be viewed, to a first approximation, as a group of interacting domains. This construct is completely general, with the domains being defined simply as the minimal cooperative units. Using this construct, we wish to examine whether the coupling between two domains in an allosteric protein can be affected by changes in the stability of a third domain. To accommodate this scenario, an allosteric protein will be represented as three domains that are allowed to interact with one another (Fig. 2-1). Because we wish to explore the role of conformational heterogeneity in allostery, each domain will be allowed the freedom to independently exist in either a low affinity tense (T) state or a high affinity relaxed (R) state, resulting in 8 possible macrostates (i.e., TTT, RTT, TRT, TTR, TRR, RTR, RRT, and RRR), representing all possible combinations of having each domain either in T or R. In the case of ID proteins, the T state would correspond to the unfolded (U) state and the R would correspond to the folded (F) state for each domain. The important feature of this model, which is the reason why each domain can "sense" the other domains, is that the free energy of any macrostate relative to the reference state (i.e., RRR) is composed of the free energy of undergoing an R to T transition in any domain (ΔG_i), plus the free energy of breaking the interaction ($\Delta g_{int,i\text{-}j}$) between the R states for the two domains in question. For this system the partition function, which is simply the sum of the statistical weights (S_i) of all macrostates, is;

$$Q = \sum_{i=1}^{N} S_i = S_{RRR} + S_{RRT} + S_{RTR} + S_{TRR} + S_{TTR} + S_{TRT} + S_{RTT} + S_{TTT} \qquad \text{(Eq. 2-1A)}$$

which, when written in expanded form using the expressions in Figure 2-1B and using the native state as the reference state, gives:

$$Q = 1 + K_3\phi_{23}\phi_{13} + K_2\phi_{12}\phi_{23} + K_1\phi_{12}\phi_{13} + K_1K_2\phi_{12}\phi_{23}\phi_{13} + K_1K_3\phi_{12}\phi_{23}\phi_{13} + K_2K_3\phi_{12}\phi_{23}\phi_{13} + K_1K_2K_3\phi_{12}\phi_{23}\phi_{13} \quad \text{(Eq. 2-1B)}$$

Examination of this expression reveals that coupling between domains results when there is a non-zero interaction energy between domains i and j (i.e., $\Delta g_{int,i\text{-}j} \neq 0$ or $\varphi_{int,i\text{-}j} \neq 1$). When the interaction energy is positive (i.e., $\Delta g_{int,i\text{-}j} > 0$ or $\varphi_{int,i\text{-}j} < 1$), it is unfavorable to break the interaction between the R states of each domain. This situation would exist, for example, when hydrophobic surfaces are excluded from interacting with solvent. In the case where $\Delta g_{int,i\text{-}j} < 0$ (i.e., $\varphi_{int,i\text{-}j} > 1$), it is energetically favorable to break the interaction and to have the surfaces interact with solvent. In the case where $\Delta g_{int,i\text{-}j} = 0$ (i.e., $\varphi_{int,i\text{-}j} = 1$), the domains are thermodynamically independent and behave as if they are simply tethered. As discussed below, the magnitude and the sign of the interaction energies significantly affects the apparent coupling between domains.

To explore the extent of coupling, perturbations to each domain can be made and the effect of the perturbation on the other domains can be monitored. For the model being investigated here, we are interested in the allosteric response of one domain to the binding of a ligand in one or both of the other domains. To accommodate this scenario, we introduce binding sites for three different ligands such that domains I, II and III in Figure 2-1 are depicted as having binding sites for ligands A, B and C, respectively. We are interested in the extent to which the binding of ligand A to domain I can influence the binding of ligand C to domain III, and we wish to know to what extent does binding of ligand B to domain II influence the observed coupling between domains I and III.

Straightforward linkage principles [82] dictate that the introduction of a ligand will result in preferential binding to the macrostate with the highest affinity, and thus a redistribution of the ensemble to favor those states with high affinity. To capture this effect, the binding of the respective ligands in Figure 2-1 is facilitated through the R state for each domain, which is

considered to be the high affinity state. Thus, macrostates RRR, RTT, RRT, and RTR can bind ligand A, macrostates RRR, TRT, RRT, and TRR can bind ligand B, and macrostates RRR, TTR, RTR, and TRR can bind ligand C, as in each case the domain that binds the specific ligand in question is in the R state (Fig. 2-1B). In the case of binding to ligand A, the partition function in the presence of ligand becomes;

$$Q_{w/A} = Z_{Lig,A}\left(S^{*}_{RRR} + S_{RRT} + S^{*}_{RTR} + S_{RTT}\right) + S^{*}_{TTR} + S_{TRT} + S^{*}_{TRR} + S_{TTT} \qquad \text{(Eq. 2-2)}$$

where $Z_{Lig,A} = (1+K_{a,A}[A])$, and $K_{a,A}$ is the intrinsic association constant of the R state of domain I for ligand A. Inspection of equation 2-2 reveals that macrostates that are competent to bind ligand A are stabilized by $\Delta g_{Lig,A} = -RT\ln(Z_{Lig,A}) = -RT\ln(1+K_{a,A}[A])$. Of course at $[A] << 1/K_{a,A}$, equation 2-2 reduces to equation 2-1. Highlighted with an asterisk (*) in equation 2-2 are those macrostates wherein domain III is also in the R state, and thus can bind ligand C. Of note is that only two of those macrostates (i.e., RRR and RTR) are stabilized by ligand A through direct binding. To determine the impact of ligand A on the ability of domain III to bind ligand C, we need only compare expressions for the probability of domain III to be in the R state, both with and without ligand A;

without ligand A;

$$P_{III,R}\left([A]=0\right) = \frac{S^{*}_{RRR} + S^{*}_{RTR} + S^{*}_{TRR} + S^{*}_{TTR}}{S^{*}_{RRR} + S_{RRT} + S^{*}_{RTR} + S^{*}_{TRR} + S^{*}_{TTR} + S_{TRT} + S_{RTT} + S_{TTT}} \qquad \text{(Eq. 2-3)}$$

with ligand A;

$$P_{III,R}\left([A]>0\right) = \frac{Z_{Lig,A}\left(S^{*}_{RRR} + S^{*}_{RTR}\right) + S^{*}_{TRR} + S^{*}_{TTR}}{Z_{Lig,A}\left(S^{*}_{RRR} + S_{RRT} + S^{*}_{RTR} + S_{RTT}\right) + S^{*}_{TRR} + S_{TRT} + S^{*}_{TTR} + S_{TTT}} \qquad \text{(Eq. 2-4)}$$

As the expressions reveal, the effect of ligand A on the probability of macrostates that can bind ligand C will depend on the statistical weights of the different macrostates in the ensemble, which

in turn will depend on the intrinsic stabilities of each domain and the interaction energies between domains (see Fig. 2-1B). To quantify the effect, we previously introduced the coupling response [41], which is a measure of the sensitivity of the population of macrostates that can bind ligand C to the addition of ligand A;

$$CR_{III,A} = \frac{\Delta P_{III,R}}{\Delta \ln Z_{Lig,A}} = \frac{P_{III,R}([A]>0) - P_{III,R}([A]=0)}{\ln Z_{Lig,A}} \quad \text{(Eq. 2-5)}$$

where $CR_{III,A}$ is the sensitivity of domain III to the binding of ligand A. The sign of the $CR_{III,A}$ determines whether ligand A is an agonist or an antagonist; if $CR_{III,A}$ is positive, ligand A is an agonist because addition of ligand A will stabilize the high affinity R state of domain III and thus increase the affinity for ligand C. Conversely, if $CR_{III,A}$ is negative, ligand A is an antagonist. In actuality, most parameter combinations produce minimal coupling between domains. However, there are multiple parameter sets that result in significant coupling that is either agonistic (Fig. 2-1C) or antagonistic (Fig. 2-1D) [41], and the hallmark feature of these parameter combinations is that the coupling is maximized when one or both domains are poised at the thermodynamic transition between the T and R states of the different domains. In the case of ID proteins, wherein the T state is unfolded, the equilibrium is poised such that one or a number of domains are significantly (=> 95%) disordered (15).

For the current study, we are interested in the impact of the binding of a second ligand (to a separate domain) on the coupling response, as described above. To account for this effect, we need only introduce the impact of the binding of ligand B (to domain II) on the probability of domain III to be in the R state, both with and without ligand A. (similar to Eqs. 2-3 and 2-4);

with ligand B and without ligand A;

$$P_{III,R}([A]=0\,|\,[B]>0) = \frac{Z_{Lig,B}(S^*_{RRR}+S^*_{TRR})+S^*_{RTR}+S^*_{TTR}}{Z_{Lig,B}(S^*_{RRR}+S_{RRT}+S^*_{TRR}+S_{TRT})+S^*_{RTR}+S_{RTT}+S^*_{TTR}+S_{TTT}} \quad \text{(Eq. 2-6)}$$

with ligands A and B;

$$P_{III,R}([A]>0\,|\,[B]>0)=\frac{Z_{Lig,A}Z_{Lig,B}S^{*}_{RRR}+Z_{Lig,A}S^{*}_{RTR}+Z_{Lig,B}S^{*}_{TRR}+S^{*}_{TTR}}{Z_{Lig,A}Z_{Lig,B}(S^{*}_{RRR}+S_{RRT})+Z_{Lig,A}(S^{*}_{RTR}+S_{RTT})+Z_{Lig,B}(S^{*}_{TRR}+S_{TRT})+S^{*}_{TTR}+S_{TTT}} \quad \text{(Eq. 2-7)}$$

The coupling response (Eq. 2-5) in the presence of ligand B thus becomes;

$$CR_{III,A}([B]>0)=\frac{\Delta P_{III,R}([B]>0)}{\Delta \ln Z_{Lig,A}}=\frac{P_{III,R}([A]>0\,|\,[B]>0)-P_{III,R}([A]=0\,|\,[B]>0)}{\ln Z_{Lig,A}} \quad \text{(Eq. 2-8)}$$

which, like Eq. 2-5, provides a measure of how the binding of ligand A to domain I influences the probability of domain III to be in the R state, except that in this case ligand B is also present. We wish to know whether ligand B can convert ligand A from an agonist to an antagonist (or vice versa).

Exhaustive search of thermodynamic parameter space

Values from -8.0 kcal/mol to 8.0 kcal/mol for ΔG_1, ΔG_2, ΔG_3, Δg_{12}, Δg_{23}, and Δg_{13} were systematically sampled in 0.1 kcal/mol increments to generate thermodynamic architectures. Values for $\Delta g_{Lig,A}$ and $\Delta g_{Lig,B}$ were set to -5.0kcal/mol and not varied as previous analyses indicated qualitative conclusions are invariant with free energy of ligand binding (See Table 2-1). All statistical weights were calculated at T = 298.15K. Parameter combinations producing agonism/antagonism switching behavior of at least ±20% were saved for subsequent analysis. All simulations were scripted in C and data was analyzed using Python and Mathematica.

2.4 Results and Discussion

Results of the Model

Transcription factors, and indeed most allosteric proteins, are considered to be either positive or negative regulators of the functions they control. It therefore might be expected that parameter combinations that are numerically close in value (i.e., stabilities and interaction energies are similar), would exhibit the same phenomenological response, being either agonistic or antagonistic, but not both. Interestingly, such a conclusion is not borne out of the current analysis. Shown in Figure 2-2 is one example of quantitatively identical parameter combinations that nonetheless produce opposite allosteric effects upon addition of ligand. For the parameter combinations noted, the energy landscape of the ensemble in the absence of ligand B is depicted in Figure 2-2A. As shown, the ensemble is dominated by the completely T state (i.e., TTT) prior to the addition of ligand A (Fig. 2-2A, right). Upon addition of ligand A, macrostates with domain I in the R state are stabilized by the amount $\Delta g_{Lig,A}$ (Figure 2-2A, left, red bars). This redistributes the ensemble, shifting the populations of all states such that there is an increase in the probability of domain III to be in the R state (Figure 2-2A, right). This leads to a macroscopic agonistic response to ligand A ($\Delta P_{III,R}$ = 51% - 3% = +48%).

The converse situation occurs in Figure 2-2B, which depicts the energy landscape of the ensemble in the presence of ligand B. The macrostates with domain II in the R state are now lower in energy relative to the situation without ligand B (Figure 2-2A, left), leaving the ensemble poised to respond in a different manner (Figure 2-2B, right). In this case, the addition of ligand A will still stabilize states with domain I in the R state as before (Figure 2-2B, left, red bars). In this case, however, redistribution reveals an antagonistic relationship between ligand A and the R state of domain III ($\Delta P_{III,R}$ = 58% - 93% = -35%, Figure 2-2B, right). The results are summarized in Figure 2-2C, where the dependence of the allosteric response on the stabilities of domains I and II is shown (all other parameters are constant). Clearly visible is that within a

range of stabilities for domain I, the sign of the allosteric response to ligand A can be modulated simply by tuning the stability of domain II. For the case examined here, the tuning of the stability of domain II is facilitated through the binding of a ligand. In reality however, the conversion can also be facilitated by mutations, covalent modification, or changes in pH. The importance of the result is that it demonstrates how a single thermodynamic architecture, within the framework of the most simple three domain model, can poise the energy landscape to respond in a "functionally pluripotent" manner. All that is required to transform an agonist to an antagonist (or vice-versa) is the presence of a second ligand that binds to a distinct regulatory site. We note that although the results presented here were generated using only one value for the binding energies for Ligands A and B (i.e., $\Delta g_{Lig,A} = \Delta g_{Lig,B} = -RT\ln(1+K_a[X]) = -5.0$ kcal/mol), varying either or both parameters does not qualitatively change the results as demonstrated below.

Table 2-1 shows $\Delta P_{III,R}(A>0)$ dependence of the free energy of ligand binding using the thermodynamic architecture from Figure 2-2 ($\Delta G_1 = -6.75$, $\Delta G_2 = -4.4$, $\Delta G_3 = -2.7$, $\Delta g_{12} = 6.8$, $\Delta g_{23} = 4.8$, and $\Delta g_{13} = -1.9$ all in kcal/mol). The bottom row (i.e. $\Delta g_{Lig,B} = 0$) is the response elicited by ligand A when ligand B is not present. Of note is that all responses are agonistic independent of ligand binding free energy of A. As the ligand binding energy of B is increased to comparable quantities (moving from the bottom row up), the ability to switch from agonistic response to antagonistic response occurs in all possible scenarios. This result demonstrates that the ability to switch responses in context of the 3 domain model from Figure 2-1 is qualitatively encoded in the thermodynamic architecture and not dependent on the free energy of ligand binding. In any case, the biological significance of the free energy does warrant consideration. Because $\Delta g_{Lig,A} = -RT\ln(1+K_{a,A}[A])$, even a ligand binding free energy of -5.0 kcal/mol corresponds to a ~1000-fold excess of ligand with respect to the dissociation constant ($K_{d,A} = 1/K_{a,A}$). As such, binding equilibria with a K_d in the picomolar to nanomolar range need ligand to be present only in the nanomolar to micromolar range respectively, concentrations that are clearly of physiological relevance.

Thermodynamic basis for agonism/antagonism switching

To determine the generality of the agonism-antagonism switching result shown in Figure 2-2, and to investigate the determinants of the switching, we performed an unbiased search of parameter space by systematically exploring all possible combinations of values for ΔG_I, ΔG_{II}, ΔG_{III}, $\Delta g_{int,I-II}$, $\Delta g_{int,I-III}$, and $\Delta g_{int,II-III}$ that produced such results. Surprisingly, parameter combinations that produced agonism/antagonism switching were highly degenerate. The stability of any particular domain or interaction energy was not critical to ensure switching potential (Fig. 2-3). Nonetheless, closer inspection of the data reveals that the organizing principles for agonism/antagonism switching center on the sign of the interaction energies between the domains. Shown in Figure 2-4A is a volume plot of the interaction energies ($\Delta g_{int,i-j}$) showing the parameter combinations that produce optimum agonism/antagonism switching. Of note is that there are four nodes of parameter combinations (Fig. 2-4A, blue). Remarkably, this plot reveals that the thermodynamic "ground rules" conferring the ability to switch responses, within the framework of the three-domain ensemble model, have two simple interaction architectures: either one or all three interaction energies ($\Delta g_{int,i-j}$) must be negative (Note: -++ corresponds to $\Delta g_{int,I-II} < 0$, $\Delta g_{int,II-III} > 0$ and $\Delta g_{int,I-III} > 0$). In contrast, the nodes that represent either one or all three interaction energies ($\Delta g_{int,i-j}$) being positive (Fig. 2-4A, gray) are not competent to switch.

Inspection of the ensemble thermodynamic architecture reveals the underlying basis for why some parameter combinations are competent to facilitate switching while others are not. For switching-competent architectures, the stabilization of domain I through the binding of ligand A will have the net effect of destabilizing domains for which the interaction energy is negative and stabilizing domains for which interaction energy is positive. For instance, consider Node 1 in Figure 2-4A, wherein, domains I and III are negatively coupled, but the other coupling energies are positive. Addition of ligand A will stabilize domain I, which will antagonistically affect domain III in a manner similar to the example in Figure 2-1D. However, since the two remaining two interaction energies are positive, stabilization of domain I will also agonistically affect

domain II, which in turn, will agonistically affect domain III. As a result, the binding of ligand A has two different and opposing effects on domain III. The overall observed effect that is manifested at domain III is the result of an energetic competition between the 'direct' interactions of domains I and III (Fig. 2-4B, Node 1, red arrow) and the 'indirect' interactions of domain I and III (i.e., those which are mediated through domain II) (Fig. 2-4B, Node 1, blue arrows). Depending on the magnitudes of the positive and negative effects, different responses can prevail. Addition of ligand B simply changes the relative contribution of the 'indirect' impact relative to the 'direct' impact. Interestingly, in some cases addition of ligand B converts an agonist to an antagonist, while in other cases, ligand B converts an antagonist to an agonist.

In contrast to the switch-competent architectures (Fig. 2-4B), the parameter combinations that result in switch-incompetent architectures (Fig. 2-4A, gray) do not produce opposing energetic effects through the 'direct' and the 'indirect' interactions. In these cases, ligand A is either a committed agonist (Fig. 2-4C, left) or a committed antagonist (Fig. 2-4C, right) because in both cases the 'direct' impact of ligand A is the same sign as the 'indirect' impact. Addition of ligand B simply reinforces the energetic relationship that existed in the absence of ligand B. Thus, the ability for allosteric proteins to facilitate agonism/antagonism switching is robustly encoded in the energy landscape, being possible even in the simplest possible architectures that can account for binding at three sites. This point is especially noteworthy considering that allosteric effects in most systems are often sensitive to changes in a second ligand (e.g., pH, salt, etc), indicating that they possess, in principle, the minimum level of complexity required to potentiate switching.

Further insight into the origins of the four nodes shown in Figure 2-4A can be gained by recasting the energetic parameter combinations in terms of the probability of domains I and II to be in the R state in the absence of ligand (i.e., $P_{I,R}$ and $P_{II,R}$). Shown in Figure 2-5 are the parameter combinations that produce $\Delta P_{III,R}$ values in excess of +/- 20% (yellow), +/- 30% (orange) and +/- 40% (red). Several features stand out. First, there are two regions that maximize

the switching behavior, and these regions correspond to cases where either one or both of the regulatory domains (i.e., domains I and II) are populating the T state a significant fraction of time in the absence of ligands (Fig. 2-5, dashed boxes). In the case of ID proteins, this scenario would correspond to being intrinsically disordered a substantial fraction of the time. Second, there is an asymmetry in the plot. Although Region 1 shows a more degenerate set of combinations of $P_{I,R}$ and $P_{II,R}$ that can facilitate the significant switching potentials (i.e. the spread in the points is greater around Region 1 than Region 2), the point density in Region 2 is three-fold higher for architectures wherein two domains are disordered (Table 2-2). For ID proteins this result would mean that switching-competent proteins are perhaps more likely to have two different disordered domains. Put another way, sequence segments that are identified as ID may actually be comprised of multiple, thermodynamically (and/or functionally) distinct, domains. Although speculative, this notion is supported by the observed differential activity of glucocorticoid receptor isoforms that differ in the length of their disordered N-terminal domains [79]. In fact, this is directly demonstrated in Chapter 5. In any case, the analysis clearly shows that for systems, which are competent to switch between agonism and antagonism, the ensemble equilibria must be poised at a point where the system is maximally able to respond to both ligands.

Implications of the model

The results presented in Figures 2 through 5 are significant not just because they demonstrate that agonism/antagonism switching can be robustly encoded in the energy landscape of a protein. They also directly undermine the (visually appealing) notion of a quantifiable allosteric pathway that mediates coupling between sites. As demonstrated by Figure 2-4A, the overall allosteric effect is a manifestation of the precise balance of energies in the protein, energies that can be modulated (in analog fashion) through the titration with ligands, salts or pH.

While it is clear that the energies from 'direct' (i.e., domain I to domain III) and 'indirect' (i.e., domain I to domain III, via domain II) contributions to the coupling should be experimentally attainable for real systems (an endeavor that, based on the results presented here, would be a logical course of action), the notion that the resultant energies can be ascribed to structural pathways between those domains is unfounded. Because the coupling is determined by intrinsic stabilities and interaction energies between each domain, any mutation that perturbs the stability will impact the coupling, regardless of whether there are changes to the interior of (or the path through) the protein. For example, it has been demonstrated that surface Val to Gly mutations, which perturb the local conformational equilibrium, can affect allosteric coupling, even in the absence of changes to the average structure [83, 84]. Such experimental results are important because they validate the tenets of the model presented here, and help to reconcile the observations that allosteric coupling; 1) can occur in the absence of a pathway of structural and dynamic changes linking the two sites [67, 70], 2) can occur in the absence of average structural changes at all [57, 58, 68, 83], and 3) can be correlated to stability changes in different parts of the protein [85]. Such observations are difficult to reconcile in terms of allosteric pathways through the protein, but are readily reconcilable in the context of the model presented here.

Relationship to classical models of allostery

The allosteric model presented here and previously [41] differs from the classic MWC [13] and KNF [80] models. The fundamental premise of MWC is that the different sites are coupled through two quaternary conformations, both of which can bind ligand. The apparent cooperativity in MWC results from a ligand driven shift from the low affinity T state to the high affinity R state. Importantly, the MWC formulation explicitly treats neither the interactions between domains nor the individual conformational equilibria describing the relative populations of T and R for each domain. The KNF model, on the other hand, does explicitly consider the

interaction between domains, but that interaction is obligatorily coupled to the binding of the ligand. As such, like the MWC, the impact of the pre-existing local conformational equilibria is not considered in the KNF model. This is the quintessential difference between the ensemble allosteric model (EAM) [41, 86] as defined here and the classic MWC and KNF formulations. By explicitly considering the (microscopic) conformational equilibria within the different functional domains, as well as the coupling of those processes, the EAM provides information that neither the MWC nor the KNF models provide - the underlying thermodynamic ground rules that relate allosteric coupling to the local energetic substructures in the protein. The importance of these ground rules is that they provide a framework for understanding the complex interplay between structure, dynamics (i.e., conformational heterogeneity) and function, a framework that provides general insight into 'how allostery works'.

2.5 Conclusions

The results presented here reveal a new dimension to allostery. Ligands for allosteric proteins have typically been viewed as either agonists or antagonists for a given function, depending on whether the activity is enhanced or repressed. Anecdotal evidence has emerged, however, that has called into question this classic view. Some allosteric effectors have been shown to be agonistic under some conditions, and antagonistic under others [77-79], although the molecular basis of this duality is not known. The results presented here reveal that even in the context of the most simple three domain architecture, protein ensembles can evolve with the capacity to respond to a ligand both agonistically and antagonistically, with the determining factor being how much of a second allosteric effector is present. In essence, rather than being considered as molecular switches with fixed responses to a particular ligand, many allosteric proteins may in fact have evolved to be "functionally pluripotent" with the ability to up or down regulate activity depending on the physiological circumstances.

Interestingly, and perhaps counter-intuitively, agonism/antagonism switching does not require that the nature of the interaction between the ligand and its binding site be different. Indeed, in the context of the model presented here, all binding (whether it produces agonism or antagonism) is facilitated by the R state of the domain (which corresponds to an equivalence of structure), a notion that is difficult to reconcile in terms of models that rely on a mechanical pathway to understand allostery. Instead, the determining factor for whether binding produces agonism or antagonism is the stabilities of the macrostates in the ensemble (which is determined by the sign and magnitude of the domain stabilities and coupling energies). Perhaps the most surprising result, and the reason why agonism/antagonism switching is even possible, is the observation that systems that are close in energetic parameter space can produce opposite phenomenological responses. This reality, as well as the significant degeneracy of potential parameter combinations that can facilitate switching, provides allosteric proteins with a robust repertoire of potential regulatory strategies.

The potential benefits of *pluripotent allostery*, as defined here, are compelling and perhaps can explain why, for instance, transcription factors "turn on" various target genes to different degrees [87]. Can the binding to a subset of DNA sequences actually repress transcription of some genes? The importance of answering this question and others like it is not just academic. It is well known that dis-regulation of many members of the SHR family of transcription factors (e.g., estrogen, androgen, progesterone and glucocorticoid receptors) are associated with cancers in various tissues [88]. Understanding how elevated or depressed concentrations, splicing variants, isoforms and/or mutations affect the allosteric communication in SHRs may be the key to deciphering their role in cancer, and to developing appropriate intervention strategies.

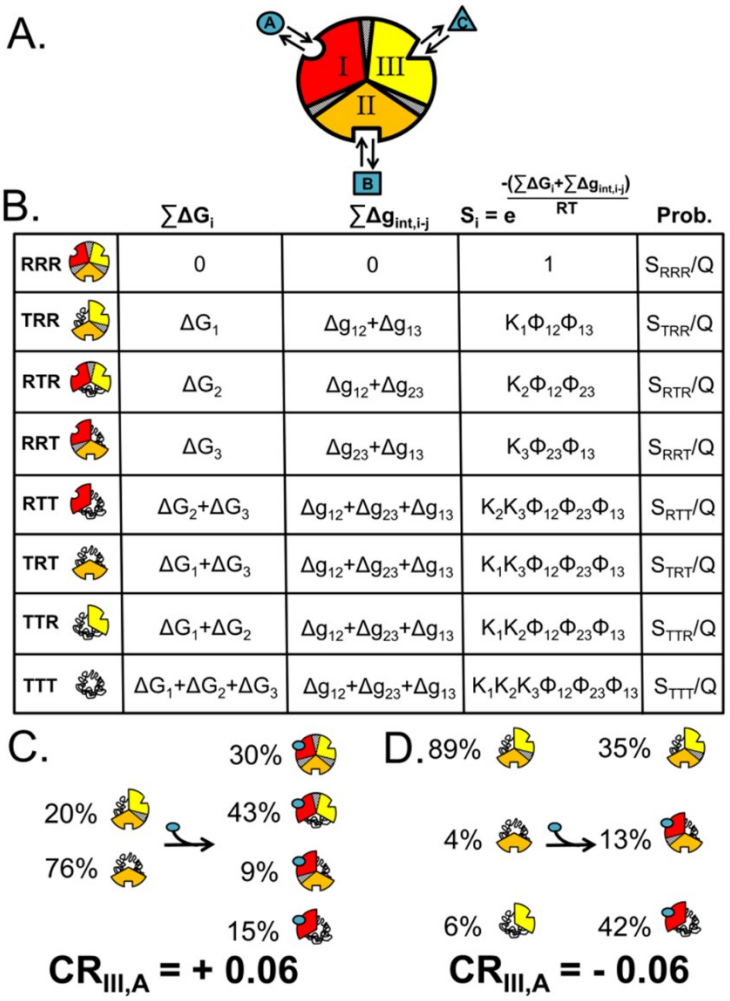

Figure 2-1. Three-domain allosteric protein. **A**. Schematic representation of a three domain allosteric protein comprised of domains I, II, and III that can bind ligands A, B, and C, respectively. Each domain is coupled to the other two domains through an interaction energy ($\Delta g_{int,i-j}$), which is schematically represented as a shaded area between domains. **B.** Free energy contributions, Boltzmann statistical (S_i) weights, and probabilities of each microstate in the ensemble (Note: $K_i=Exp[-\Delta G_i/RT]$ and $\varphi_{i-j}=Exp[-\Delta g_{i-j}/RT]$). Each macrostate represented in the left column is either completely high-affinity (RRR), partially high-affinity (TRR, RTR, RRT, RTT, TRT, TTR), or completely low-affinity (TTT). The free energy contributions come from the transition of each

domain to the low-affinity T state ($\sum\Delta G_i$) plus the energy of breaking the interactions between domains ($\sum\Delta g_{int,i-j}$). **C.** A specific case demonstrating an agonistic allosteric response. Shown are states that are significantly populated (i.e., >3%) before (left) and after (right) the addition of ligand A. In the absence of ligand A, the summed probability of macrostates with domain III in the R state (i.e., the active states – domain III colored yellow) are only marginally populated (~20%). Upon addition of ligand A, redistribution of the ensemble results in a positive shift in the population of these states (~73% = 43% + 30%), which corresponds to an agonistic response (i.e.,$CR_{III,A}$ = +0.06). The parameters used are ΔG_1 = -1.7, ΔG_2 = 2.0, ΔG_3 = -0.9, Δg_{12} = -2.3, Δg_{23} = 0.1, Δg_{13} = 1.5, and $\Delta g_{Lig,A}$ = -5.0 all in kcal/mol. **D.** A specific case demonstrating an antagonistic allosteric response. Unlike the case of agonism, the active states are significantly populated in the absence of ligand A (~95% = 89% + 6%). Upon addition of ligand A, redistribution of the ensemble results in a negative shift in the population of these states (35%), which corresponds to an antagonistic response (i.e.,$CR_{III,A}$ = -0.06). The parameters used are ΔG_1 = -2.1, ΔG_2 = 1.0, ΔG_3 = 1.2, Δg_{12} = -1.7, Δg_{23} = 0.6, Δg_{13} = -2.7, and $\Delta g_{Lig,A}$ = -5.0, all in kcal/mol.

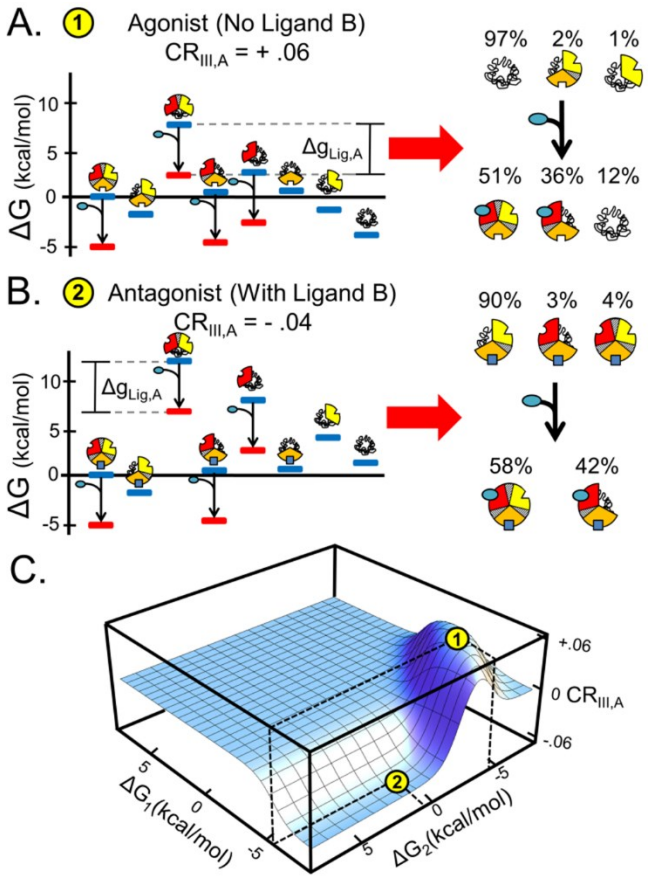

Figure 2-2. Specific thermodynamic architectures can produce agonism/antagonism switching. Example of a thermodynamic architecture that produces agonism-antagonism switching: ΔG_1 = -6.75, $\Delta G_{2,B=0}(1)$ = -4.4, $\Delta G_{2,B>0}(2)$ = 0.6, ΔG_3 = -2.7, Δg_{12} = 6.8, Δg_{23} = 4.8, Δg_{13} = -1.9, and $\Delta g_{Lig,A}$ = -5.0 kcal/mol. Individual free energies and populations for each microstate showing an **A.** agonistic and **B.** antagonistic response. Shown for both cases (left) are the free energies of each state in the absence (blue bars) and presence (red bars) of ligand A (depicted as blue circles). Shown also for each case are the populations of states in the absence (right, top) and the presence (right, bottom) of ligand A. For the case of agonism (**A**), the summed population of macrostates with domain III in the R states rises from 3% to 51%, while for antagonism (**B)** those states decrease from 94% to 58%. **C.** Coupling response (CR) dependence on stability of domains I and II

(holding all other variables constant). Position 1 (yellow circle at local maximum) exhibits agonistic response to ligand A in the absence of ligand B. Position 2 (yellow circle at local minimum) exhibits antagonistic response to ligand A in the presence of ligand B ($\Delta g_{Lig,B}$ = -5.0 kcal/mol).

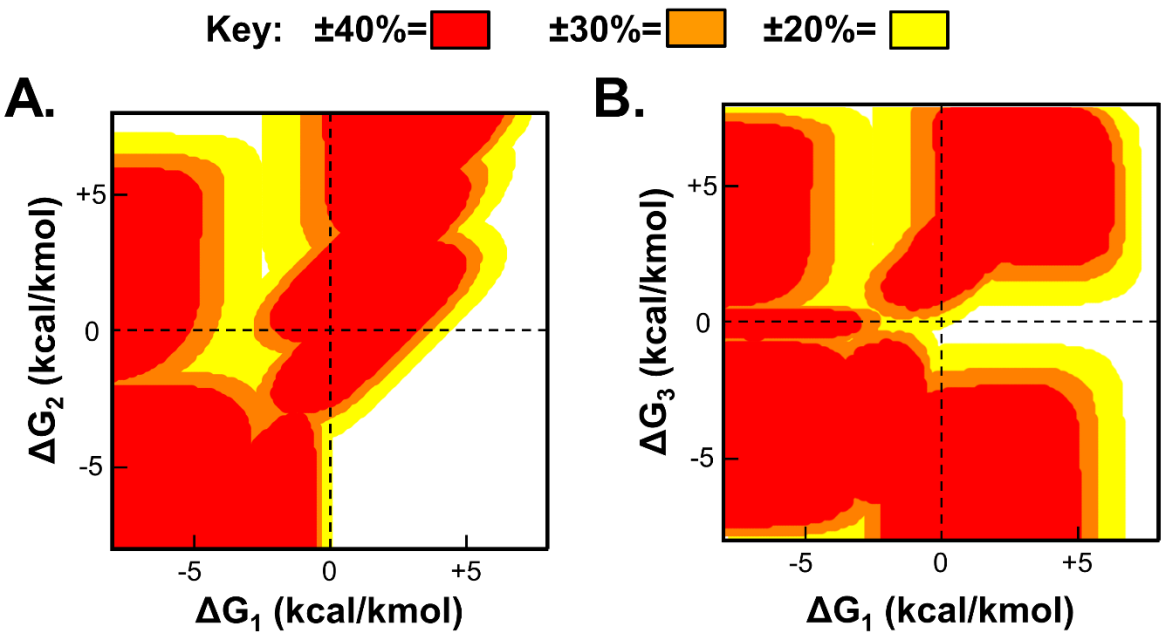

Figure 2-3. Agonism/Antagonism switching is highly degenerate for stabilities of domains. Exhaustive search of thermodynamic parameter space reveals that the free energies of domains 1, 2, and 3 (i.e. ΔG_1, ΔG_2 and ΔG_3 respectively) are highly degenerate for switching behavior to the same ligand. **A.** Plot of ΔG_1 versus ΔG_2 values that resulted in agonism/antagonism switching architectures with thresholds of ±40% (red), ±30% (orange), and ±20% (yellow). These values correspond to CR thresholds of ±40%, ±30%, and ±20% respectively. **B.** Plot of ΔG_1 versus ΔG_3 values that resulted in agonism/antagonism switching architectures with thresholds of ±40% (red), ±30% (orange), and ±20% (yellow). Both **A** and **B** contain clusters of parameters that maximize switching corresponding to the four nodes depicted in Figure 2-4A.

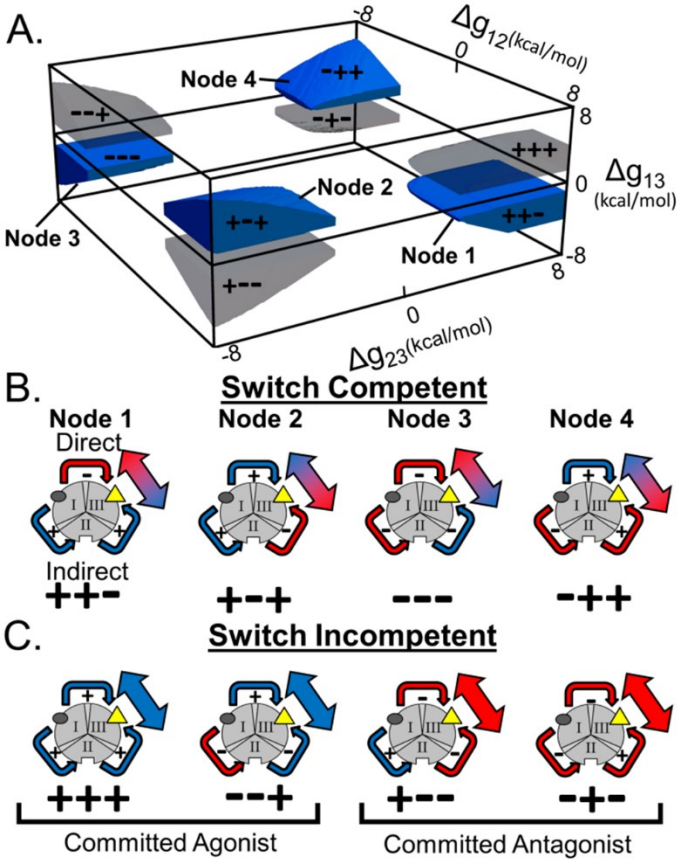

Figure 2-4. Energetic rules for agonism-antagonism switching. A. Plot of interaction parameters (i.e. Δg_{12}, Δg_{23} and Δg_{13}) that exhibit significant agonistic (>35%) and antagonistic (< -35%) responses within a single architecture (blue nodes labeled 1-4). Each node is represented by three + or – signs that correspond to the sign of the coupling energies, Δg_{12}, Δg_{23}, and Δg_{13}, respectively. For contrast, parameter combinations that do not exhibit switching are shaded in gray (See text for details). **B.** Schematic representation of the allosteric interdomain coupling for each of the switch-competent nodes in (A) with each node numbered and represented by three + or -. Red and blue arrows indicate an antagonistic and an agonistic relationship between two domains, respectively. The red and blue arrow indicates that domain III is impacted both positively and negatively through the 'direct' and 'indirect' effects, as described in text.

Note: The color of the arrows connecting domains II and III refer to the overall impact on domain III from stabilization of domain I. For example, in Node 3 stabilization of domain I destabilizes domain II (making the arrow red). However, that destabilization of domain II (because of negative coupling to domain III) has the effect of stabilizing domain III (making the arrow to domain III blue). **C.** Schematic representation of the allosteric interdomain coupling for each of the switch-incompetent nodes in (A). The coloring scheme is similar to B and demonstrates that switch-incompetent architectures produce either committed agonists (+++, --+) or committed antagonists (+--, -+-) because the 'direct' and 'indirect' effects are of the same sign.

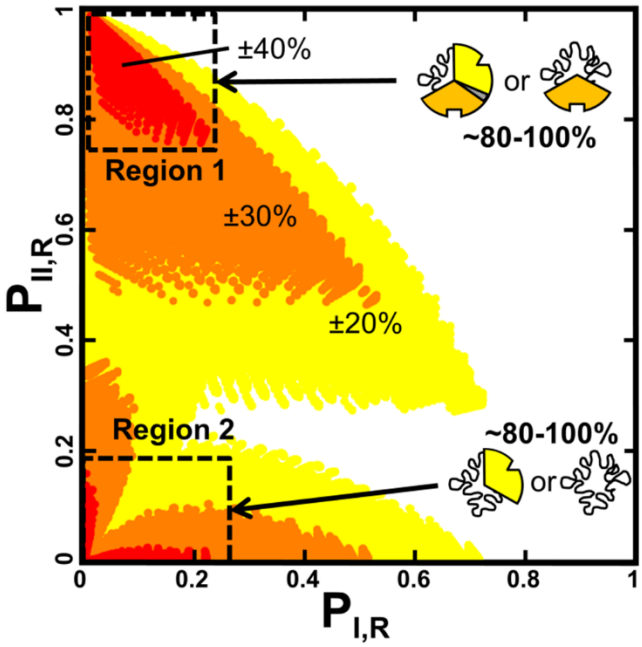

Figure 2-5. Switching-competence is maximized when regulatory domains are predominantly populating low affinity or ID states. Plot of the probability of domain I being in the R state (i.e. $P_{I,R} = P_{RRR}+P_{RTR}+P_{RRT}+P_{RTT}$) *vs.* probability of domain II being in the R state (i.e. $P_{II,R} = P_{RRR}+P_{TRR}+P_{RRT}+P_{TRT}$) in the absence of ligand for switching-competent architectures. Colors indicate the magnitude of the change in probability for agonist and antagonist switching: yellow (±20%), orange (±30%), and red (±40%), i.e. the architecture must exhibit an agonistic and an antagonistic response from ligand A exceeding the positive and negative thresholds in the absence and presence of ligand B (or vice-versa). Maximum switching responses (dashed boxes) are observed in two regions, where either domain I only (i.e. Region 1) or both domains I and II (i.e. Region 2) are predominantly populating the T state (or in the case of ID proteins, the unfolded (U) state).

Table 2-1 - Agonism/Antagonism switching behavior is encoded in the thermodynamic architecture and not qualitatively dependent on ligand binding free energy.

$\Delta g_{Lig,B}$ (kcal/mol) \ $\Delta g_{Lig,A}$ (kcal/mol)	-5.0	-4.0	-3.0	-2.0	-1.0
-5.0	-35.2%	-34.7%	-32.3%	-23.2%	-8.1%
-4.0	-31.5%	-31.0%	-28.8%	-20.4%	-7.0%
-3.0	-15.6%	-15.4%	-14.0%	-9.4%	-3.0%
-2.0	20.0%	19.4%	16.4%	8.8%	2.2%
-1.0	45.8%	40.7%	25.4%	8.2%	1.5%
0.0	48.0%	31.0%	10.5%	2.2%	0.4%

Table 2-2. Point density of Regions 1 and 2 show Region 2 is more often populated.

Threshold	±30%		±35%		±40%	
	Ratio[b]	Percentage[c]	Ratio	Percentage	Ratio	Percentage
Region 1[a]	1	27.90%	1	26.90%	1	24.70%
Region 2[a]	2.58	72.10%	2.71	73.10%	4.05	75.30%

[a]Regions 1 and 2 are defined as points greater than and less than 40% probability of domain II being R respectively (see Figure 2-5).

[b]Ratio calculated by counting the number of points in each region normalized to Region 1.

[c]Calculated by percentage of points in respectful region compared to the total number of points meeting threshold criterion.

Chapter 3 – Single-molecule force spectroscopy instrumentation and calibration

3.1 Abstract

The observation that light momentum could be used to manipulate small particles spawned multiple fields of study the past few decades. In particular, it was realized that single biological macromolecules could be manipulated with extreme precision revealing a wealth of mechanistic information. The emergence of single-molecule force spectroscopy has allowed the measurement of forces on the order of piconewtons and extension changes on the order of nanometers for various biological processes. However, this instrumentation is not commercially available which requires custom manufacturing. Here we describe certain aspects of the MiniTweezers laser optical trap that was built and calibrated to experimentally test hypotheses concerning allostery and intrinsic disorder. The instrument reports on only three observables: time, distance, and force. Distance and force are measured by position sensitive detectors that directly measure photon flux while time is measured by a computer's internal clock. Taking advantage of the accurate actuator motors of the microfluidics chamber reveals a robust method to relate photon flux to force and distance. Calibration factors are tested by mechanically manipulating single molecules of DNA revealing that it is routinely possible to calibrate the instrument to within 1% of the actual values. Furthermore, we experimentally test the 2nd law of thermodynamics by directly measuring entropy production from Brownian motion and are able to directly observe spontaneous negative entropy processes as predicted by the fluctuation theorem. These results lend credence to the proper construction and calibration of the MiniTweezers that were used in all subsequent experiments.

3.2 Introduction

The realization that light could be used to trapped and manipulate small particles of different dielectrics spawned multiple fields in the latter half of the past century [89, 90]. The application of optical traps to biological macromolecules has been a more recent development from the past two decades [91]. Studies initially focused on measuring the physical properties of biopolymers such as DNA [92] and the detailed mechanism of molecular motors [93, 94]. Advances in the chemistry of labeling polystyrene beads, eventually led to the first single-molecule force-spectroscopy folding experiments on DNA [95], RNA [96], and proteins [97]. These single-molecule folding studies revealed rich information on biologically meaningful intermediates and transient states that would have been otherwise obscured in bulk, directly demonstrating the power of this technique specifically, and of single-molecule experiments in general.

Towards to goal of performing our own single-molecule folding studies to dissect allostery and coupling within intrinsically disordered proteins, we built our own optical MiniTweezers instrument in collaboration with Dr. Dmitri Toptigyn (Johns Hopkins University) and Dr. Steve Smith (University of California Berkeley). These instruments are not commercially available, and thus, need to be built by hand. In particular, this instrument reports on forces on the order of piconewtons and distances on the order of nanometers. As a result, the measurements are sensitive to multiple variables including but not limited to: high pressure weather systems, footsteps within the room, and even a human heart beat. As such, reliable measurements requires precise manufacturing, calibration, and maintenance.

This chapter outlines the optical elements within the instrument that were aligned (Figures 3-1 and 3-2), and discusses the theoretical basis of converting raw voltages measured from the instrument to real experimental values. We calibrated the instrument's force detection using the hydrodynamic drag of a polystyrene bead with known radius, and the distance by taking

advantage of the highly precise actuator motors that move the microfluidics chamber [98]. We tested the accuracy of calibration by pulling on single-molecules of DNA and found that the measurements are accurate to within 1% of their true values. Furthermore, we pushed the limits of detection of the instrument by measuring "violations" of the 2nd law of thermodynamics consistent with what has been previously reported [99] – i.e. over small time scales we are able to directly measure negative entropy events from thermal noise as predicted by the fluctuation theorem [100]. These data lend credence to the proper construction of the instrument and lead us to conclude that all data collected on the instrument in subsequent chapters are exceptionally precise and reliable.

3.3 Materials and Methods

Optical MiniTweezers instrumentation

The optical MiniTweezers system was built in collaboration with Dr. Dmitri Toptygin (Johns Hopkins University) and Dr. Steve Smith (University of California Berkeley) along with instructions available online (http://tweezerslab.unipr.it/). All electronic components were purchased from their respective manufacturers listed on the website, metal parts were ordered from eMachinesShop (http://www.emachineshop.com/), and printed circuit boards were soldered by Dr. Dmitri Toptygin. A general overview of the instrument setup is presented in Figure 3-1. Initial optical alignment was performed as described in the online directions. We built the instrument using 845nm telecom lasers that are manipulated using a piezo setup dubbed the "fiber wiggler" (green bars, top of Fig. 3-2). The fiber wiggler is a patented technology by Dr. Carlos Bustamante and Dr. Steve Smith (US Patent: 7274451) that allows precise movements of the trap position on the order of nanometers. Furthermore, the fiber wiggler increases stability of the

entire system since the microfluidics chamber is held stationary allowing for this instrumentation to be used in normal lab space as opposed to sub-basements that are on separate foundations.

Single-molecule DNA handle generation

DNA tethers were generated using two different methods: non-covalent tethers that utilizes both biotin/streptavidin (bio/SA) and digoxigenin/anti- digoxigenin (dig/anti-dig), and one that covalently links a 5' end of the handle to the bead and uses a bio/SA on the other end. For both methods, DNA handles were amplified from the Lambda phage genome using the following primers - forward: 5'-ACTGATGCAACTGACTCAGC-3' and reverse: 5'-TTGGATATCAGAGCTATGGC-3'. These primers amplify a 2.97kbp (herein referred to as 3kbp) region of the Lambda phage genome that is ~50% GC and uniform in GC content as calculated by a moving average. Product was confirmed by agarose gel electrophoresis.

For non-covalent DNA handles, the forward primer has a 5'-dig moiety and the reverse primer has a 5'-bio moiety. This region was amplified by PCR, the product was run on an agarose gel, purified by means of a Thermo Scientific gel-extraction kit, and then isopropanol precipitated. Concentration and purity were confirmed by absorption spectroscopy. To generate tethers, two types of beads were required: one coated with SA and another coated with anti-dig. 2.1μM SA beads were purchased from Spherotech while anti-dig beads were synthesized based on a variation of a previously described method [96]. 2.8μM carboxylated beads were purchased from Spherotech as well for anti-dig modification. These beads were pelleted and resuspended in activation buffer (100mM MES, 0.5M NaCl, 1% Tween20, pH 6.0), and mixed with fresh ~5mM EDC (1-Ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride) and ~10mM sulfo-NHS (N-hydroxysuccinimide) resuspended in deionized water. Beads were stirred at room temperature for 15 min, quenched with 1mM 2-mercaptoethanol, and washed with activation buffer. At this

point, protein G (Pierce, 21193) was added the beads and tumbled at room temperature for 3 hours allowing crosslinking to occur. Beads were then spun down, washed with 1x PBS, and mixed with Sheep Polyclononal Anti-Dig Antibody (Roche 1333-089) and 10mM DMP (dimethyl pimelimidate). The second crosslinking reaction was allowed to proceed for two hours at room temperature. Beads were finally spun down, washed in 1x PBS with 10% glycerol, and flash frozen with liquid nitrogen. For covalent DNA handles, the forward primer with the dig moiety was replaced by one that had a primary amine. This was covalently linked to the carboxylated beads using the EDC/sulfo-NHS protocol described above for protein G coupling.

Single-molecule tethers of DNA were generated *in situ* (i.e. in the microfluidics chamber). A streptavidin bead was trapped, taken to the micropipette tip, and held by suction. A second bead with either covalent or non-covalent DNA handles was then trapped and taken to the micropipette tip. Tethers were generated by moving the fiber wiggler's piezo and bringing the two beads within close physical proximity. Tethers were confirmed by force generation and the curvature of the entropic elasticity. Single-molecules were confirmed from the overstretching transition that occurred at ~65pN. All DNA pulling experiments were conducted in 10mM Tris, 0.5M NaCl, 1mM EDTA, pH 7.5.

Direct measurement of entropy from thermal noise

Polystyrene beads (3.4μm) were purchased from Spherotech and exchanged to deionized water via centrifugation and pipetting. Beads (5% w/v) were diluted 10,000:1 and injected into the microfluidics chamber. Once a bead was trapped, the system was precisely aligned by adjusting the light lever mirrors, moving the fluidics stage by hand, and minimizing the force signal via the piezos that engage the fiber wigglers. At this point, laser power was reduced to ~1% of its maximum power (~0.5mW) and the system was allowed to equilibrate for a few

minutes. The bead was then moved at a constant velocity of 100nm/s and all data were recorded at 1kHz. Analysis and theory of these data are described in the results and discussion section below.

3.4 Results and Discussion

Force Calibration

A general overview of the optical path of the laser light is presented in Figure 3-2. The instrument has a counter propagating dual laser beam single trap as previously described [98]. We will take advantage of the properties of photons to calibrate the instrument. Light photons carry a momentum **P** such that:

$$\boldsymbol{P} = \hbar \boldsymbol{k} \qquad \text{(Eq. 3-1)}$$

where $\hbar$ is reduced Plank's constant and **k** is the light wave vector. The physical basis of trapping polystyrene beads depends on the refractive index difference between the aqueous media the bead [89, 90]. By refracting the photons, any trapped object will experience a momentum change that is equal but opposite of the change in light momentum change as described by Eq. 3-1. Position Sensitive Detectors (PSDs) at multiple locations in Figure 3-2 directly measure the flux of photons along the optical path. As a result, we have access to the net photon flux, and therefore, the net momentum and force that a particle experiences in the optical trap.

An external force on a trapped bead will manifest itself by a deviation in the mean position of the photon flux (Figure 3-3) yielding a change in position (Δx). This a result of the angular deflection of the bead which we will denote θ. Assuming the bead is significantly larger than the wavelength of light we can use a simple ray optics approach as opposed to a quantum Maxwell equation approach. Using Snell's law and assuming the laser light only interacts with

air outside of the chamber, we can find a relationship between the displacement Δx and the angle of deflection such that:

$$\frac{\Delta x}{R_L} = n_L \sin\theta \qquad \text{(Eq. 3-2)}$$

where R_L is the focal length of the lens, n_L is the refractive index of the liquid media, and all other variables are as previously defined. The transverse light force generated by deflection of a ray of light is given by:

$$F = \left(\frac{n_L W}{c}\right) \sin\theta \qquad \text{(Eq. 3-3)}$$

where F is the force experiences, W is the intensity of the ray (i.e. photon flux), c is the speed of light, and θ is as previously defined. By combining Eqs. 3-2 and 3-3 the transverse light force that the bead experiences under an external force becomes:

$$F = \left(\frac{W}{c}\right)\left(\frac{\Delta x}{R_L}\right) \qquad \text{(Eq. 3-4)}$$

where F is the force experienced, W is the light ray intensity, R_L is the focal length of the lens, and Δx is as described.

Two main conclusions for calibration purposes can be drawn from Eq. 3-4. The first is that there is a direct relationship between force and the change in position of the photon flux Δx. The second is that the proportionality between force and Δx depends on photon flux which can be directly measured by a PSD. Most optical traps are calibrated by considering this direct relationship and treating the trap as an effective "spring" such that:

$$F = k * \Delta x \qquad \text{(Eq. 3-5)}$$

where k is the Hooke spring constant and all other variables are as previously defined. Indeed, this spring constant is related to W and is in fact the calibration constant that we wish to obtain for our trap to directly measure forces [98].

The technique we used to measure our force calibration constants was by viscous drag of a polystyrene bead in liquid media. This is known as a, "Stokes' Law," calibration which relates the viscous drag of a spherical particle to other physical properties:

$$F_d = 6\pi\,\mu\,R\,v \qquad \text{(Eq. 3-6)}$$

where F_d is the drag force, μ is the dynamic viscosity of the liquid media, R is the radius of the spherical particle, and v is the velocity. We have access to every variable in Eq 3-6 which we can then use to relate the raw voltages measured by the force PSDs to the exact force. We have access to dynamic viscosity because we will use water at a known temperature [101], the radius of the particles based on scanning electron microscopy and flow cytometry of the polystyrene beads, and the velocity of the microfluidics chamber because of the actuator motors. It is of note that the actuator motors that move the microfluidics chamber can measure distances on the order of microns with great precision (used for distance calibration below), and the combination of a computer's internal clock yields an accurate measurement of chamber velocity.

Performing a Stokes' law calibration involves trapping a bead, moving the microfluidics chamber in the X, Y, and Z directions independently, and measuring the raw voltages from the PSDs as a function of motor velocity. Representative data are demonstrated in Figure 3-4A for the x-direction. This plot reveals two important features. The first is that the voltage in the X-dimension (yellow) is linear with motor velocity in the X-direction supporting the assertion that the relationship should be linear from Eq. 3-4. Second, and perhaps more importantly, the Y and Z PSD voltages (red and black respectively) do not have any relationship with the motor velocity

in the X-direction. This demonstrates that the PSDs have been properly aligned and placed inside the instrument otherwise we would observe non-linear behavior.

Repeating the experiment described above for the Y and Z directions, we can use linear regression to obtain the slope of the line between PSD voltage and motor velocity, and subsequently calculate the k-calibration constant based on Eq. 3-6. Indeed, trapping a 3.4μm polystyrene bead and moving the motors in all three directions reveals that this methodology yields calibration values that are within 1% of the true expected values (Figure 3-4B). In this figure, the bead was moved at room temperature (22.94ºC according to the thermal detectors in the instrument) which yields an expected slope of 0.029pN*s/μm denoted by the dashed blue line. The values obtained from the instrument all fall within experimental error of the expected value (red, green, and blue lines). The validity of these calibration constants are also verified by pulling on single-molecules of DNA (see below). In short, this method yields calibration values that support previous theory and can be routinely performed.

Distance Calibration

Distance calibration also takes advantage of the actuator motors. However, it also uses a constant-force fast-feedback protocol of the MiniTweezers which is why force calibration is always performed beforehand. The relative trap position is measured using PSDs similar to force and the signal comes from approximately 1% of the laser light is picked off by a pellicle beam splitter before it enters the microfluidics chambers. As a result, the measured voltage by the PSDs report on the relative position that the fiber wiggler is aiming the trap in the microfluidics chamber (Figure 3-2). The position PSDs are commonly referred to as “Light Levers” and also take advantage of the linear range of response from the PSDs.

In the calibration experiment, a single bead is trapped and held by the micropipette tip in the middle of the microfluidics chamber. The trap is then placed over the bead trying to generate as little force as possible (<10pN). The instrument is then put into a constant-force fast-feedback protocol that attempts to maintain the same force, and therefore the same position. The motorized stage is then moved a finite distance and the piezos on the fiber wiggler move an equal but opposite distance to maintain the constant force. This direct relationship between the piezo movement and chamber movement can be used to calibrate the Light Lever signal. The true calibration values are the ones that create this unity relationship between the known and measured distance change.

Representative data from such an experiment is shown in Figure 3-5. It is of note that there is a clear linear relationship between the Light Lever and motor position which again supports the notion of a linear relationship between displacement and voltage detected by the PSDs. What also stands out from this plot is that are regions where the light-lever values do not change while the motor values do as can be appreciated by the vertical lines. These are due to motor-backlash that occurs when the motor direction is changed. This is a technical issue that is unavoidable when moving motors on the micron length-scale. These portions are not used in the fitting procedure for calibration constants.

From these plots, the calibration values are determined by linear regression and calculating the value required to create a unity relationship. These final values along with the force calibration constants are shown in Table 3-1. For comparison are calibration values from an instrument from Dr. Carlos Bustmante's lab at University of California Berekeley referred to as, "Mendel". What is clear is that all of our calibration values are within 10% of the Mendel's calibration values except for the Z-force constants. This deviation in addition to all other deviations are because of microscopic differences in assembly of the instrument. The significant difference in Z-force constants is ascribed to our Z-force PSD being slightly raised since our focal

plane is slightly different. Regardless of the origin of this effect, the values are demonstrated highly accurate in the following sections.

Single-molecule force-spectroscopy on 3kbp DNA handles

To test the validity of the calibration process, we tethered single-molecules of 3kbp DNA handles between two beads (Figure 3-6A). Tethers were pulled at constant velocity (100nm/s) until a maximum force of 80pN was achieved, and then relaxed back down at the same velocity. Using this pulling method, we were able to obtain single-molecule tethers for the covalent and non-covalent DNA handles (Figure 3-6B). It was empirically noted that non-covalent tethers had a much shorter half-life (i.e. spontaneous tether "breaking" occurring from dissociation of dig/anti-dig) once force was increased above 50pN as previously reported [102].

The data in Figure 3-6B are what we would expect for a single-molecule of DNA. In the low force regime, there is some slight curvature originating from the entropic elasticity of heteropolymers [92]. Once the force has increased to ~10pN, we reach a linear regime in which the distance between the beads does not change, but the trap position is still increasing [98]. This is known as the trap stiffness regime, where the force is directly proportional to the trap position and the force is generated from the stiffness of the trap. Finally, at ~65pN we observe what is referred to as an overstretching transition. This is the hallmark of a single-molecule as this is the transition from B-form DNA to an overstretched form that is believed to remove all stacking interactions between bases and leave only hydrogen bonds between the base pairs [95].

These results indicate that our instrument is calibrated properly by two independent metrics. The first is that the overstretching occurs at exactly 65pN. Most protein folding events will occur in the 5-30pN range, and thus if the instrument is accurate at 65pN, all the data below that force regime are interpolated and thus extremely reliable. In fact, the transition in Figure 3-

6B cannot be statistically differentiated from 65pN demonstrating that we are within 1% of the true force calibration value. The second result that is that the overstretching transition corresponds to exactly the length we would expect for 3kbp of double-stranded DNA as we see an increase in contour length of ~1.7 times once overstretching is complete [95]. Taken together these results indicate that our instrument has been built properly, is able to perform pulling experiments, and is properly calibrated.

Experimental "violations" of the 2nd law of thermodynamics

The 2nd law of thermodynamics can be stated in various forms, but the form most applicable to this discussion is that every spontaneous process will not occur with 100% efficiency. That is, it is impossible to spontaneously get useful work out of the surroundings into the system. Taught even at the high-school level, this law implies that the entropy or "randomness" of any spontaneous event must increase – i.e. $\Delta S > 0$ for all real processes [101]. One of the greatest intellectual achievements of statistical thermodynamics during the past few decades was the fluctuation theorem (FT) that addresses this law in a rigorous statistical formalism [100]. The FT formalized how it is possible for events to occur where the spontaneous entropy was negative, an apparent "violation" of the 2nd law. Of course this is not a true violation, it is simply stating the statistical nature of entropy and was even first pointed out by Boltzmann. Since entropy is an extensive property, its magnitude scales with size of the system. At a macroscopic level in which bulk experiments are performed, we are trapped in a regime where entropy must be greater than zero. However, optical trapping provides an interesting opportunity. We are able to observe thermal noise on the order of milliseconds and at the length scale of nanometers. This is right near the threshold at which it is possible to directly measure such "violations" of the 2nd law.

The FT formalizes the relationship between a trajectory over observation time t producing entropy ($S_t = A$) versus the same trajectory producing the same entropy in magnitude, but opposite in time:

$$\frac{P[S_t=A]}{P[S_t=-A]} = \exp(A) \qquad \text{(Eq. 3-7)}$$

Since entropy is an extensive property of the system, S_t will scale with the dimensions of the system. This is truly an astonishing relationship and cannot be overstated. Eq. 3-7 states that there is a finite probability that the *surroundings will spontaneously put work into the system*.

In the experiment described in the methods section, we will test to see if our instrumentation is sensitive enough to measure some of these finite trajectories described by Eq. 3-7. From Eq. 3-5 we know that we can treat displacement as a Hookean spring. Thus, if we consider a beads trajectory in time x(t) and the optical forces acting on this bead, we can calculate the entropy production (St) during this time:

$$S_t = \frac{1}{k_B T} \int_0^t F(s)\, v(s) ds \qquad \text{(Eq. 3-8)}$$

where k_B is the Boltzmann constant, T is the absolute temperature, and F(s) and v(s) are the force and velocity over the trajectory integrated from time 0 to time t [99].

For most processes, we expect to observe Figure 3-7A where the water molecules will be randomly colliding with the bead that is trapped, and will not produce any efficient work on the system being the bead. However, as Eq. 3-7 suggests, there is a finite probability that we will observe a random alignment of water molecules as depicted in Figure 3-7B, where the surroundings will spontaneously put work into the system, i.e. a negative entropy process. Similar to previous results, we were unable to detect any negative entropy production when we numerically integrated Eq. 3-8 on the order of seconds [99]. However, when we integrated on the order of 10ms we are directly observe these "violations" of the 2nd law (Figure 3-7C). At first

glance, it does not seem that there are any negative entropy productions. However, the inset in Fig. 3-7C directly demonstrates that an extremely small portion of events indeed reveal "violations" of the 2nd law of thermodynamics.

There are some differences in the shape of the distribution from our data and those previously published [99]. These differences are mostly contributed to a change in bead size and a different experimental setup in terms of the optical trap (e.g. MiniTweezers versus a trap setup on an optics table). Regardless of the origin of this effect, the conclusion is clear. The MiniTweezers instrument that was built is able to be reliably calibrated, can reproduce experimental data that has already been published, and is sensitive enough to directly observe negative entropy processes.

3.5 Conclusions

Here we have presented a brief overview of how the optical MiniTweezers system was built and a method to robustly calibrate the instrument. Our results indicate that despite the shortcoming of Stokes' law calibration [98], we are able to obtain calibration constants within 1% of their true values. Furthermore, the distance calibration is exceptionally accurate as demonstrated by the measurements in contour length of DNA and measurements made on protein unfolding in the following chapter. Indeed, the instrument is able to even measure negative entropy processes even when the lasers are turned down to ~0.5mW, demonstrating the remarkable sensitivity of the instrument. Perhaps most importantly, this method of calibration provides a routine maintenance method that was employed over the course of the last few years. Should there be any misalignment of the instrument, the calibration method and experiments mentioned in this chapter served a benchmark to detect optical aberrations before permanent damage to the instrument was incurred. In short, these results indicate that the MiniTweezers

instrument was built properly, calibrated, and will be able to perform new experiments to measure unknown properties of allosteric and intrinsically disordered proteins.

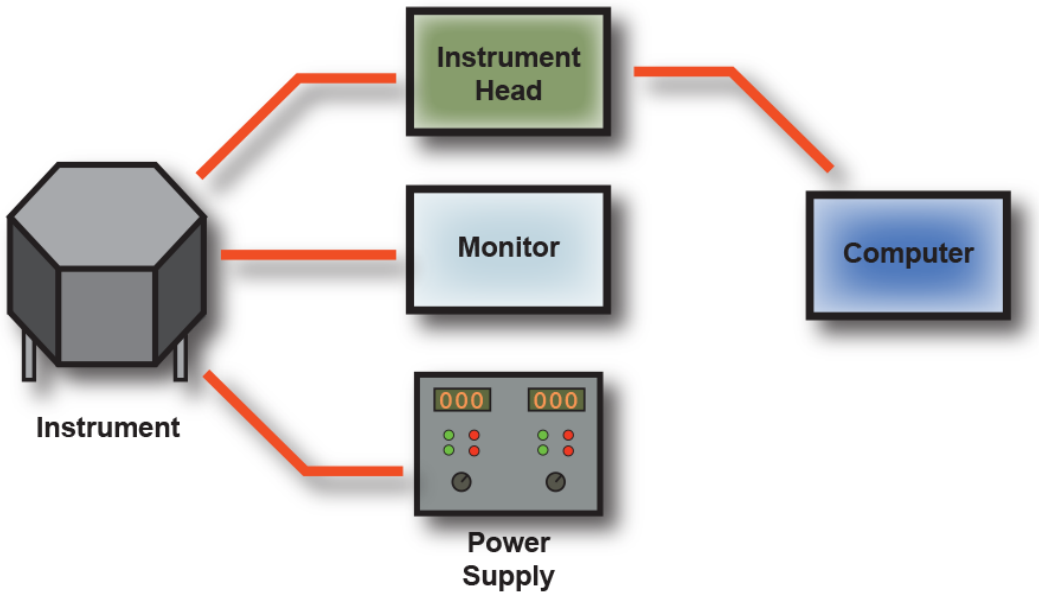

Figure 3-1. General overview of the optical MiniTweezers instrument. Shown are the five main components of the optical MiniTweezers. The main instrument (where the experiment is conducted) is shown on the far left. The instrument is able to communicate to three separate components shown in the middle: (1) the power supply for the lasers, (2) the television monitor that allows the user to see the beads that they are trying to trap, and (3) the instrument head. The instrument head contains the printed circuit boards that have soldered microchips with experimental procedures for the fast feedback protocols involved in constant-force and constant-velocity experiments. The computer is connected to the instrument head via a USB cable that allows the user to instruct the instrument on the specifics of how the experiment will be conducted. All data is sent from the instrument to the instrument head where it is averaged to 1kHz and then sent to the computer for storage.

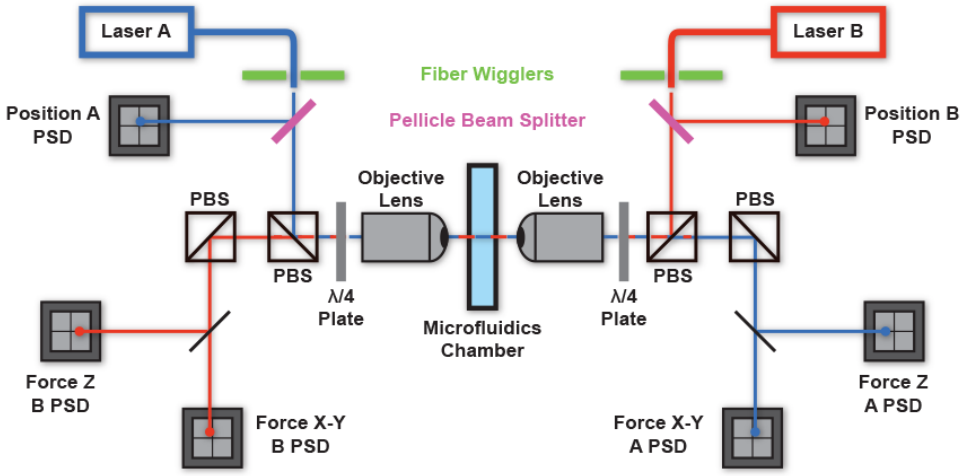

Figure 3-2. Laser optical path for the optical MiniTweezers. Shown is a coarse schematic of the optical setup inside of the instrument. The instrument consists of two lasers: A (blue) and B (red) shown at the top. The quartz crystal end of the laser (~5µm wide) is manipulated by the fiber wiggler (green bars, see text for details). As laser light leaves the wiggler, ~1% of the light is picked off by a pellicle beam-splitter (purple bars) to the position detectors (PSD). These report on the relative extension change based on where the lasers are aimed inside the microfluidics chamber. The majority of the light goes to a photo-selective beam splitter (PBS) and through a quarter wavelength plate (λ/4) which converts the polarity of the light from linear to circular. Once going through the objective lens, the light is focused inside the microfluidics chamber where the trap is made. Upon exiting the chamber, light enters the far objective lens, has its polarization converted again, and eventually relays off of a final PBS to the force detectors. The force PSDs report on how much the light has been refracted by the particle in the microfluidics chamber. Forces for X and Y are measured separately from Z for technical purposes beyond the scope of this thesis. The optical path is identical for both traps except that they are counter-propagating.

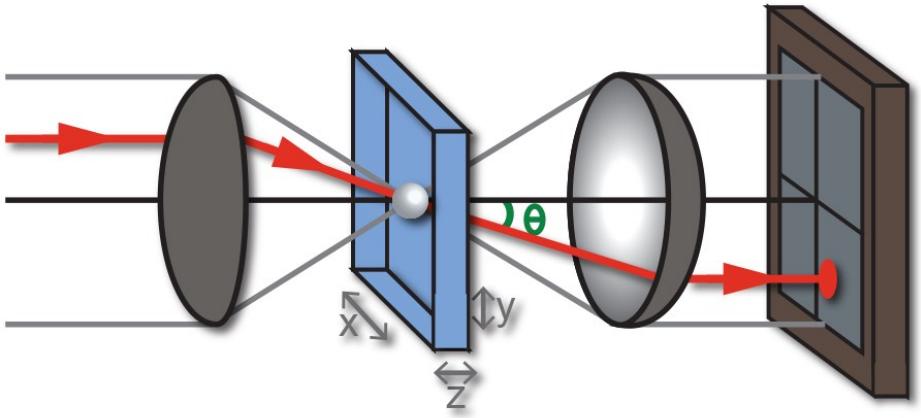

Figure 3-3. Force detection depends on the angle of incident light. Simplified schematic of laser path in the optical tweezers setup. The laser light is moving from left to right through two objective lenses (gray semi-circles) and a microfluidics chamber (blue box). When the polystyrene bead inside the microfluidics chamber experiences force, the laser light leaving the chamber is refracted and exits at angle θ. The laser light finally hits the position sensitive device (PSD) on the far right and its displacement from the center and distribution determines the X, Y, and Z forces.

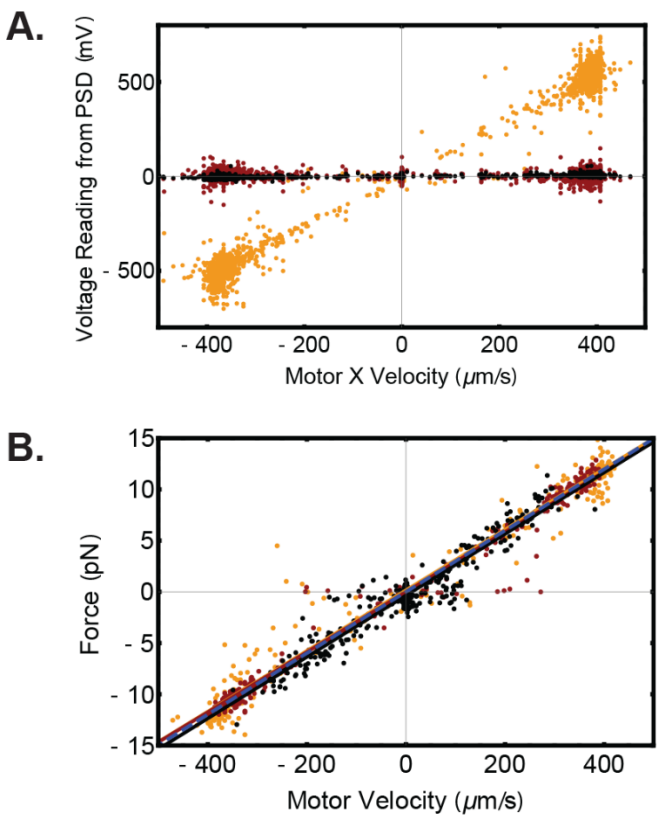

Figure 3-4. Force calibration representative data. **A.** Shown are the results from a Stokes' law calibration for the x-direction. In yellow are the X-PSD values, and red and black show the Y-PSD and Z-PSD values respectively. There is a clear linear relationship between the motor velocity (μm/s) and the voltage reading (mV) as predicted by the discussion in the text. Additionally, the lack of relationship with Y and Z demonstrate that the PSDs have been properly aligned in the instrument. **B.** Shown are the results of Stokes' law calibration for X, Y, and Z in yellow, red, and black respectively. The dashed blue line shows the expected line for a bead of 3.4μm at room temperature. All fits are within 1% of the expected value indicating that the force calibration is complete.

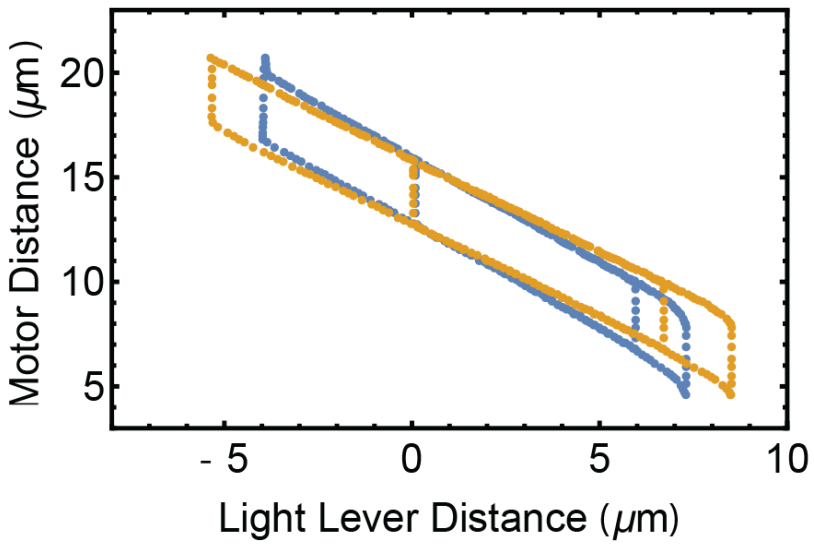

Figure 3-5. Light lever calibration representative data. Depicted are the light-lever readings for the X-direction for the A-laser (blue) and B-laser (yellow). There is a clear linear relationship between the motor and light-lever distance. Of note is that there are regions where the motor distance changes while the light levers do not (vertical regions). These are due to motor backlash and are an unavoidable technical challenge. Also of note is that the slopes of the A-laser and B-laser are different, suggesting that the instrument is not fully calibrated. The slopes of these lines are used to establish a unity relationship, and thus, the calibration constants for the light-levers.

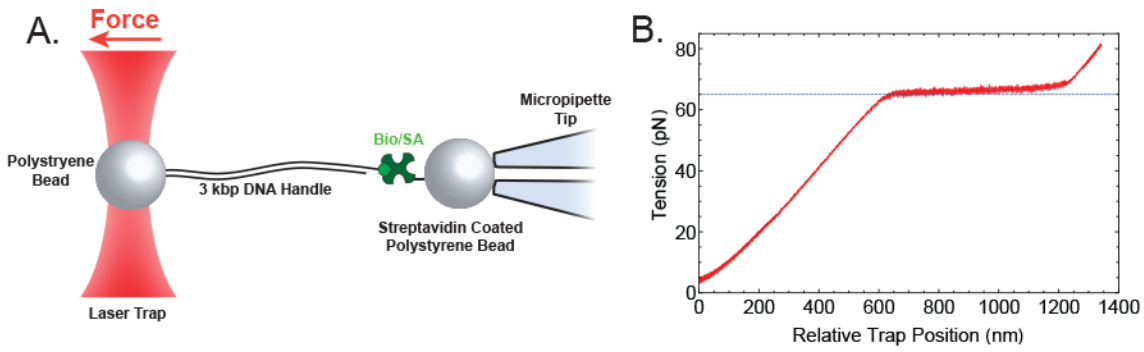

Figure 3-6. Single-molecule force spectroscopy on DNA handles reveals proper calibration of instrumentation. A. Depicted is the single-molecule experimental setup to pull on double-stranded DNA molecules. A bead is held in the optical trap which is covalently linked to 3kbp of DNA. The 5' end of the DNA handle has a biotin moiety which is non-covalently tethered to a streptavidin coated bead that is held in the micropipette tip by suction. Force is generated by moving the optical trap at constant velocity. **B.** Representative data from the single-molecule experiment described. Shown is tension (pN) versus relative trap position (nm). Note that the overstretching transition occurs exactly at 65pN (blue dashed line) indicating that the force is properly calibrated. Furthermore, the contour length change and distance of the overstretching transition are consistent with expected values indicating that the distance calibration is complete as well.

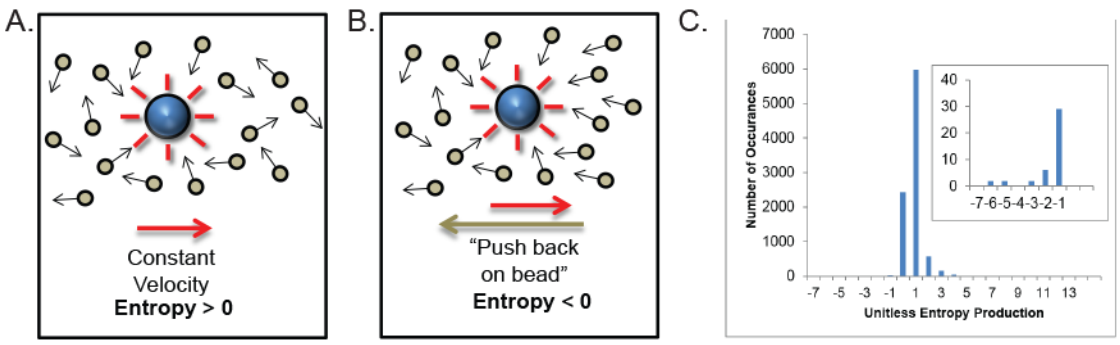

Figure 3-7. Experimental "violations" of the second law of thermodynamics. A. Depicted is a positive entropy process where the bead is randomly colliding with water molecules (gray circles) as it is moved at a constant velocity (red arrow, pointing right). **B.** Depicted is a spontaneous negative entropy process where enough of the water molecules align to push back on the bead (gray arrow, pointing left) over small enough time scales. This is an example of when work is spontaneously put into the system, an apparent "violation" of the second law of thermodynamics. **C.** Experimental data from performing such an experiment. Plotted is the number of occurrences (n) versus unitless entropy production as described in the text. The vast majority of traces result in zero or positive entropy production. The inset shows a zoom in of the area with negative entropy. What is clear is that a small fraction of traces generate negative entropy processes.

Table 3-1. Calibration constants for optical MiniTweezers instruments demonstrate proper calibration.

	Hilser Lab MiniTweezers			Mendel MiniTweezers		
	X	Y	Z	X	Y	Z
TrapA Force Sensitivity	0.01331	0.0125	0.02826	0.01496	0.01465	0.012
TrapB Force Sensitivity	0.01263	0.01286	0.02826	0.0144	0.01433	0.012
TrapA Lever Sensitivity	9313	8154	0	9025	7966	0
TrapB Lever Sensitivity	9543	7992	0	8996	8155	0

Chapter 4– Two-dimensional single-molecule force spectroscopy allows structural modeling of protein folding

This work has been submitted for publication at the writing of this thesis:

Motlagh HN*, Toptigyn D, Kaiser CM, and Hilser VJ*. (2015) Two-dimensional single-molecule force spectroscopy allows structural modeling of protein folding. Biophysical Journal, (Submitted)

*To whom correspondence should be addressed: hnekoor1@jhu.edu or hilser@jhu.edu

4.1 Abstract

Single-molecule force spectroscopy has emerged as a powerful tool for studying the folding of biological macromolecules. Mechanical manipulation has revealed a wealth of mechanistic information on transient and intermediate states. To date, the majority of state assignment of intermediates has relied on empirical demarcation. However, performing such experiments in the presence of osmolytes provides an alternative approach that reports on the structural properties of intermediates. Here, we analyze the folding and unfolding of T4 lysozyme with optical tweezers under a chemo-mechanical perturbation by adding osmolyte. We find that two unrelated protective osmolytes, trimethylamine-N-oxide (TMAO) and sorbitol, function by marginally decelerating unfolding rates and specifically modulating early events in the folding process, stabilizing formation of an on-pathway intermediate. The chemo-mechanical perturbation provides access to two independent metrics of the relevant states during folding trajectories, the contour length and solvent accessible surface area. We demonstrate that the osmolyte dependence of the intermediate, in conjunction with its measured contour length, provides the ability to discriminate between potential structural models of intermediate states. Our study represents a general strategy that may be employed in structural modeling of equilibrium and intermediate states observed in single-molecule experiments.

4.2 Introduction

Single-molecule force spectroscopy has emerged as a powerful tool for studying protein folding [97, 103, 104]. Devoid of ensemble averaging, mechanical manipulation of single protein molecules with optical tweezers has revealed a wealth of mechanistic information on transient states and intermediates that would have otherwise been obscured in bulk measurements[46-48, 96, 97, 103-108]. To date, state assignment of these intermediates and structural transitions has relied on changes in contour length and empirical domain boundary demarcation [48, 96, 107, 108]. Only recently, studies have begun to identify structural intermediates through mutational analysis[47], although such experiments are time consuming, difficult, and not applicable to all systems. Performing force spectroscopy experiments in the presence of osmolytes, however, represents an alternative approach, providing a chemo-mechanical perturbation that reports on the relative accessible surface area of states in addition to contour length [49-51, 103, 109].

Protective osmolytes are ubiquitous small organic molecules that have evolved in nature to counteract the deleterious effects of harsh environmental conditions experienced by many organisms[1]. In elasmobranchs, for example, up to 0.4M urea can accumulate in the cytosol, introducing significant denaturing stresses to the proteins in the cytosolic milieu[1]. Protective osmolytes function by stabilizing folded conformations of proteins, and despite the fact that osmolytes fall into different chemical classes (e.g., polyols and methylamines), the mechanism of stabilization of proteins appears to be similar. Osmolytes interact unfavorably with the peptide backbone, thus stabilizing the compact native state, wherein exposed backbone surface is minimized [49, 109]. A detailed understanding of how osmolytes modulate the energy landscape of proteins (known as the osmophobic effect) has proven elusive.

In particular, it is clear from bulk measurements that osmolytes affect both folding and unfolding kinetics [110-113], yet mechanistic information on how they change the folding pathway is lacking. Although single-molecule studies with atomic force microscopy (AFM) have

revealed that in high concentrations of osmolyte, unfolding kinetics are slowed and compaction of unfolded species occurs faster when force is quenched [114-117], those studies were performed in a force regime that did not permit direct observation of folding.

To address these limitations and to directly observe the effect of osmolytes on the folding of individual molecules, we performed single-molecule force spectroscopy experiments on the cysteine free variant of T4 lysozyme [T4*] [118] using optical tweezers. T4* consists of two coupled subdomains; a sequentially discontiguous N-terminal domain and a sequentially contiguous C-terminal domain (Fig. 4-1A). T4* has been extensively studied both in bulk [119, 120] and at the single-molecule level [46, 48], and is known to fold through an intermediate state. As such, it is an ideal system to perturb with osmolyte to characterize both the unfolded to intermediate, as well as the intermediate to native state transitions.

We conducted single-molecule experiments in the absence and the presence of two different osmolytes, sorbitol and trimethylamine N-oxide (TMAO), representing two different classes of osmolyte molecules (Fig. 4-1B). Their impact on the mechanical unfolding and refolding rates was determined. Because our single-molecule force spectroscopy experiments provide both the extension change as well as the osmolyte sensitivity of the intermediate state population, we are able to reliably determine the structural properties of the intermediate state. In addition, we show that we are able to discriminate between multiple detailed structural models of the intermediate. The resultant approach may represent a general strategy for characterizing structures populated during protein folding, thereby adding to the library of information that can be gleaned from single-molecule experiments.

4.3 Materials and Methods

Pulling Buffers and Materials

All pulling experiments were conducted in buffer HKM (25mM HEPES*KOH, 150mM KCl, 5mM $MgCl_2$, pH 7.4). Trimethylamine N-oxide dehydrate (TMAO) was purchased from Sigma-Aldrich and dissolved with the components of pulling buffer. To remove impurities from TMAO synthesis, all buffers were stirred with activated carbon (12-20 mesh; Sigma-Aldrich) for 1 hour while protected from light. Sorbitol was purchased from Sigma-Aldrich. The buffers were filtered (0.22um filter; Millipore), divided into aliquots, and stored at -80^{o}C until use. All beads were purchased from Spherotech.

Protein purification

The plasmid harboring the single-molecule cysteine-free variant of T4 Lysozyme (T4*) was provided from a previous study [48]. The construct contains an N-terminal Avi Tag for biotinylation [121], a flexible linker, the open reading frame of T4*, and then a YbbR tag on the C-terminus for addition of Coenzyme A (CoA) crosslinked to oligonucleotide [122]. The insert was subcloned into a pET vector with a 9X-His-Tag and a TEV protease site adjacent to the Avi-tag used for biotinylation. The plasmid was transformed into Rosetta™ 2(DE3)pLysS (EMD Millipore) and plated on LB plates with 100μg/mL ampicillin. The next day colonies were picked and grown to preparative volumes at 37^{o}C 250rpm until the OD_{600nm} reached 0.6-0.8. Expression was induced with 1mM IPTG under the same conditions for 3 hours. Cells were harvested by centrifugation and stored at -80^{o}C. Pellets were resuspended in Lysis Buffer (50mM NaH_2PO_4, 300mM NaCl, 20mM Imidazole, pH 8.0) and lysed by 5 passes through an Avestin EmulsiFlex C3 at 15,000psi. Lysates were clarified by centrifugation and immediately loaded onto HisPur Ni-NTA Superflow Agarose (Thermo Scientific) and eluted with a linear gradient

with the same buffer containing 500mM imidazole. Samples containing T4* were pooled and dialyzed to TEV digestion buffer (50mM Tris-HCl, 0.5mM EDTA, 1mM DTT, pH 8.0) along with TEV protease (1μM final concentration). Cleaved T4* was recovered by running through the same nickel affinity resin, dialyzed to 1x PBS, concentrated to 100μM, and flash frozen with liquid nitrogen.

Synthesis of Oligonucleotide crosslinked to CoA

Oligonucleotide crosslinked to CoA (OligoCoA) was synthesized as previously described [48, 106] except OligoCoA was purified by reverse-phase high-performance liquid chromatography (HPLC). After crosslinking, samples were isopropanol precipitated and resuspended in water. The samples were then desalted by an illustra MicroSpin G-25 Column (GE Healthcare) according to the manufacturer's instructions. Samples were concentrated by vacuum centrifugation and then purified by semi-preparative reverse-phase HPLC on an Agilent 1100 Series instrument. Modified oligonucleotide was separated using a 250x10-mm Phenomenex Jupiter C18 column packed with 5μm particles. Mobile phase A was 100mM Triethylammonium acetate (TEAA) pH 7.0 and mobile phase B was the same except with the addition of 15% acetonitrile. The column was equilibrated to 60^{o}C and samples were eluted by a linear gradient of 0%B to 100%B over 30 minutes and then holding 100%B for another 30 minutes while detecting absorption at 260nm. Pure fractions were confirmed by Urea-PAGE. Samples were finally isopropanol precipitated and confirmed by MALDI-TOF.

Labeling of protein for optical tweezer experiments

The labeling occurred in two major steps: first a biotinylation reaction specific to the N-terminus followed by attachment of CoA (that was crosslinked to double stranded

oligonucleotide) to the C-terminus. Prior to the optical tweezer experiments, these molecules were ligated directly to polystyrene beads. During experiments, tethers were generated via a non-covalent biotin/streptavidin/biotin interaction *in* situ (i.e. in the microfluidics chamber) between the N-terminal biotin and the streptavidin/biotin complex on the surface of the DNA bead held in the trap (see below). Many of these methods are previously described [48, 106] and modifications to these protocols are described for completeness.

The T4* construct harbored an N-terminal Avi tag sequence (GLNDIFEAQKIEWHE) which is specifically modified by *E. coli* biotin ligase (BirA) at lysine 10 in the sequence[121]. BirA was purified as previously described [48]. Purified protein was mixed with 1x biotinylation buffer (20x = 500μM D-biotin, 100mM ATP, 100mM Mg(OAc)), 2μM BirA, and allowed to react for 2h at room temperature. Samples were then dialyzed against 1x PBS to remove excess biotin. Biotinylated protein was purified by using Pierce Monomeric Avidin Resin (Thermo Scientific) according to manufacturer's instructions. Samples containing biotinylated protein were dialyzed to 1x PBS.

The T4* construct also harbored a C-terminal YbbR tag (DSLEFIASKLA) which is specifically modified at serine 2 by the phosphopantethein moiety of CoA in an enzymatic reaction catalyzed by the phosphopantethein transferase Sfp from *Bacillus subtilis[123]*. Sfp was prepared as previously described [122]. OligoCoA was annealed to a complimentary strand that had a 4bp-overhang allowing for downstream annealing and ligation of protein to beads. Biotinylated T4* (25μM) was labeled by mixing freshly annealed dsOligo-CoA (50μM), Sfp (10μM), and 1x HM buffer (5x = 50mM Hepes*KOH, 10mm $MgCl_2$, pH 7.5) and incubating for 2 hours at 37^{o}C. Modification was confirmed by mobility shift in SDS-PAGE and unused dsOligo-CoA and Sfp were removed via Pierce Monomeric Avidin Resin (Thermo Scientific). Samples were then dialyzed to 1x PBS overnight, diluted to 5μM, and flash frozen with liquid nitrogen.

Synthesis of DNA handles

DNA handles were generated by crosslinking a 5'-primer to polystyrene beads [102] and directly amplifying the handle on the bead in a PCR-like fashion. The DNA handle was ~1.8kbp with ~50% GC content and was amplified from the Lambda phage genome. Carboxylic acid coated polystyrene beads (2.8μm diameter) were spun down and washed in 100mM MES pH 6.0 multiple times to remove any residual material from manufacturing. The carboxyl groups on the beads were then activated with 40μg/μL fresh EDC (1-Ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride) three times for a 15 minute incubation period. Beads were then quickly washed in 50mM HEPES pH 7.5 and excess 5'-primer was added (5'-NH_2-ACTGATGCAACTGACTCAGC-3'). The reaction gently shook at room temperature for two hours before quenching via addition of 1M Tris, pH 7.5 to a final concentration of 10mM. The beads were then washed extensively with MilliQ water to remove any unreacted product. The beads were finally spun down and resuspended in an equal volume of 5mg/mL Casein (Affymetix) and 1% Tween-20 to prevent bead aggregation. The beads were treated as a 5'-primer in a PCR reaction with a 3'-primer harboring a biotin moiety on its 5'-end (5'-bio-GAAAGAATGGGCATGAGC-3') as previously described[48]. After PCR, the beads were spun down and washed extensively using TE Buffer with 20% glycerol and then finally flash frozen with liquid nitrogen and stored at -80^{o}C. For optical tweezer measurements, beads were thawed and incubated with streptavidin before dilution and injection into the microfluidics chamber. This protocol yielded stable tethers for multiple rounds of pulling and relaxing up to the overstretching transition at 65pN [95].

Synthesis of beads for protein ligation

Beads were synthesized as previously described [48] except different primers were used. The two primers used were: 5'-NH_2-CATGCGTCCTGATGTTAGCTCTCCG-3' and 5'-

CGCACGGAGAGCTAACATCAGGACGCATG-3' which leaves a 4bp overhang complementary to the OligoCoA/protein.

Optical tweezers measurements and analysis

The optical MiniTweezers system was built and calibrated according to directions available online (http://tweezerslab.unipr.it/). The instrument has a counter propagating dual laser beam single trap which measures force directly by the change in light momentum [98]. Experiments were conducted in a microfluidics chamber with a micropipette tip and purchased from Steve B Smith Engineering (Optical Instruments, Biophysics Consulting). Beads ligated with protein were trapped, moved to a micropipette tip, and held by suction. Subsequently, a bead with DNA handles and streptavidin was trapped and brought within close physical proximity of the protein bead in the micropipette tip. Tethers were confirmed by exerting force and observing behavior described previously [48]. Single-molecules were confirmed by overstretching of DNA handles [124]. Force and trap position were recorded at a sampling frequency of 400 Hz for all constant velocity (force ramp) and constant force (force clamp) experiments. The total number of single-molecule tethers generated under each condition was: N_{Buffer} = 32 molecules, $N_{Sorbitol}$ = 20, and N_{TMAO} = 24 molecules. Constant force rates were determined from a significant number of transitions (total transitions: N_{Buffer} = 1,173; $N_{Sorbitol}$ = 605; N_{TMAO} = 1,237).

Derivation of transfer free energy for intermediate states

The single-molecule folding traces yield direct access to the probability of the intermediate (P_I) relative to the unfolded state. All calculations below only consider the change in probability from Buffer to 1M TMAO. Since we have measured $P_{I,Buffer}$ and $P_{I,1M\ TMAO}$, we can

calculate the free energy change of the chemo-mechanical perturbation using a Boltzmann distribution as follows:

$$\frac{P_{I,1M\ TMAO}}{P_{I,Buffer}} = \frac{\exp\left(\frac{\Delta G_{I,1M\ TMAO}}{R*T}\right)}{\exp\left(\frac{\Delta G_{I,Buffer}}{R*T}\right)} = \exp\left(\frac{-[\Delta G_{I,1M\ TMAO} - \Delta G_{I,Buffer}]}{R*T}\right) \quad \text{(Eq. 4-1)}$$

where $\Delta G_{I,1M\ TMAO}$ and $\Delta G_{I,Buffer}$ are the free energies of the intermediate state relative to the unfolded state (i.e. reference state), R is the universal gas constant, and T is the absolute temperature. Implicit in this formalism is the assumption of equilibrium. We consider this a justified assumption given the low force regime (i.e. near equilibrium) and the subsequent predictive capabilities of the model generate. Should the assumption of equilibrium be incorrect, the predictive capabilities would be compromised.

The value in the numerator on the right side of Eq. 4-1 can be re-written as a function of the transfer free energy of the intermediate and unfolded states:

$$\Delta G_{I,1M\ TMAO} - \Delta G_{I,Buffer} = G_{I,1M\ TMAO} - G_{U,1M\ TMAO} - G_{I,Buffer} + G_{U,Buffer} = \Delta G_{I,tr} - \Delta G_{U,tr} \quad \text{(Eq. 4-2)}$$

Where $\Delta G_{I,tr}$ and $\Delta G_{U,tr}$ are the transfer free energies of the intermediate and unfolded state to 1M TMAO respectively. Both of these transfer free energies depend on the accessible surface area (ASA) of the states in question. As mentioned in the main text, we treat the intermediate state as a contiguously folded portion of the protein that has the same ASA as the crystallographic structure. This is justified since the intermediate is on-pathway and is presumably native-like.

Let us consider a general intermediate within the context of the derivation above: the intermediate has a contiguously folded portion of amino acids N_{tr} through C_{tr}, where these values are integers corresponding to the amino acid numbers that are the boundaries for the folded portion of the molecule (Note: $1 \leq N_{tr} < C_{tr} \leq 164$). Several immediate relationships become apparent from this formalism. For instance, the unfolded portion of the molecule is amino acids 1 through N_{tr-1} and C_{tr+1} through 164 (the number of amino acids in T4*). Since these residues are

also unfolded in the unfolded state their transfer free energies in Eq. 4-2 cancels out. The transfer free energy in equation 2 actually corresponds to the transfer free energy of the folded portion (i.e. amino acids N_{tr} through C_{tr}). This can be appreciated by calculating the free energy of both states:

$$\Delta G_{I,tr} = \sum_{i=1}^{N_{tr}-1}(\Delta\alpha_i^{SC}\Delta g_{tr,i,SC}^{o} + \Delta\alpha_i^{BB}\Delta g_{tr,i,BB}^{o}) + \sum_{j=C_{tr}+1}^{164}(\Delta\alpha_j^{SC}\Delta g_{tr,j,SC}^{o} + \Delta\alpha_j^{BB}\Delta g_{tr,j,BB}^{o}) \quad \text{(Eq. 4-3)}$$

and

$$\Delta G_{U,tr} = \sum_{i=1}^{164}(\Delta\alpha_i^{SC}\Delta g_{tr,i,SC}^{o} + \Delta\alpha_i^{BB}\Delta g_{tr,i,BB}^{o}) \quad \text{(Eq. 4-4)}$$

where i is the amino acid type from the primary sequence, Δg^{o}_{tr} is the free energy of transfer for the side-chain (SC) or backbone (BB) to 1M TMAO, and $\Delta\alpha_i$ is the fractional change in solvent ASA from the unfolded to the intermediate state [49, 50].

The $\Delta\alpha_i$ values in Eqs. 4-3 and 4-4 require estimates of the ASA of the unfolded and intermediate states:

$$\Delta\alpha_i^{bb\ or\ sc} = \frac{\sum_{j=1}^{n_i}(ASA_{i,j,U} - ASA_{i,j,I})}{n_i ASA_{i,Gly\cdot X\cdot Gly}} \quad \text{(Eq. 4-5)}$$

Where the numerator is summed over all amino acids j of type i, $ASA_{i,jU}$ is the ASA of the unfolded state, $ASA_{i,j,I}$ is the ASA of the intermediate state, n_i is the total number of groups of amino acid (AA) type i, and $ASA_{i,Gly.X.Gly}$ is the standard side-chain or backbone solvent accessibility of Gly-X-Gly tripeptides presenting the maximally exposed surface area (X it the amino acid of type i) [125]. The values for α are calculated based on the steered molecular dynamics simulations to represent the unfolded state (see text in next section).

Substituting Eqs. 4-3 and 4-4 into Eq. 4-2 results in a canceling out of terms such that now Eq. 4-2 reduces to:

$$\Delta G_{I,1M\ TMAO} - \Delta G_{I,Buffer} = -\sum_{N_{tr}}^{C_{tr}}(\Delta\alpha_i^{SC}\Delta g_{tr,i,SC}^{o} + \Delta\alpha_i^{BB}\Delta g_{tr,i,BB}^{o}) \quad \text{(Eq. 4-6)}$$

where all the terms are defined identically to Eqs. 4-3 and 4-4. The right hand side of Eq. 4-6 is actually a calculation of the transfer free energy of the folded portion of the intermediate to 1M

TMAO. In particular, Eq. 4-6 is what was used to calculate the heat map in Figure 4-8A and to relate the experimental probability changes to the contour plot (i.e. by substituting Eq. 4-6 into Eq. 4-1). In all subsequent calculations, each amino acid was treated as either folded or not based on the intermediate boundaries defined by N_{tr} and C_{tr}.

Steered molecular dynamics simulations

All simulations were performed with VMD and NAMD molecular dynamics software package [126] with the CHARMM36 force field [127]. Initial structure of T4* was obtained from the PDB (PDB ID: 1L63). Due to the large extension of the fully elongated conformation (~600Å), explicit solvent was intractable given limited computational resources. All simulations were performed using an implicit solvent model: Generalized Born (GB) with a cutoff radius of 12Å and an ion concentration of 0.15M similar to the optical tweezer experiments. Temperature was controlled using Langevin dynamics (damping coefficient 2/ps) at 300K. VDW interactions were treated as a Lennard-Jones 6-12 potential calculated using the switch method starting at 13Å and smoothly cutoff at 14Å. The SHAKE method was used on all hydrogen atoms allowing integration steps of 2fs expediting simulation time.

After hydrogens were built in, the structure was minimized by conjugate gradient for 1,000 steps. The system was then equilibrated for 4ns. Three random points were selected within a 2ns window of the end of equilibration and used as different starting points in the subsequent simulations. Steered molecular dynamics simulations were performed by fixing the C-terminal Cα and pulling on the N-terminal Cα with a spring constant of 1000pN/Å and a velocity of 4nm/s. The system was pulled until the approximate contour length of T4* was achieved (~58.1nm). Three trials of each of these simulations were performed. Trajectories were analyzed by using the VMD software and subsequent analysis scripts described below for surface area and energy calculations.

Structure based calculations for the intermediate state

The ΔΔG values for intermediates were calculated based on the group transfer free energy model which uses experimentally measured free energies of transfer for every amino acid side chain and the peptide backbone [49, 50]. Potential unfolded states were generated by mechanically unfolding T4* *in silico* (described above). All other potential unfolded state conformations were taken from the SMD simulations described in the previous section. Intermediate states were treated as having the crystal structure conformation for the folded portion and the unfolded conformation for the remainder. The accessible surface area (ASA) to a 1.4Å rolling sphere was calculated for all states using a subroutine from the COREX algorithm [128]. From there, the change in ASA from unfolded to intermediate was calculated for every possible species (i.e. every possible combination of the N-terminus and C-terminus unfolded). Using these values it was possible to calculate ΔΔG as follows:

$$\Delta\Delta G = \Delta Gtr, U - \Delta Gtr, I = \sum_{i=AA\ type}[n_i \Delta g^o_{tr,i,SC}\, \Delta\alpha_i^{SC}] + \Delta g^o_{tr,BB} \sum_{i=AA\ type}[n_i \Delta\alpha_i^{BB}] \qquad \text{(Eq. 4-7)}$$

Where n_i is the total number of groups of amino acid (AA) type i, Δg^o_{tr} is the free energy of transfer for the side-chain (SC) or backbone (BB) from 0M to 1M TMAO, and $\Delta\alpha_i$ is the fractional change in solvent ASA from the unfolded to the intermediate state [49, 50]. The $\Delta\alpha_i$ values used in the calculation above were the mean values of those consistent with the unfolded state: i.e. all states that fell in the range $\mu_U \pm \sigma_U$ = 25.3±2.0nm.

4.4 Results and Discussion

Osmolytes minimally affect the mechanical unfolding pathway

Single molecules of T4*, subjected to continuously increasing force ("force ramp" experiments), exhibited cooperative unfolding (Fig. 4-2A). The mean unfolding forces ($<F_{unf}>$) obtained in the presence and absence of osmolytes overlap significantly: $F_{unf,buffer}$ = 17.6±1.9pN (N=224), $F_{unf,1M\ Sorbitol}$ = 18.2±2.1pN (N=288), and $F_{unf,1M\ TMAO}$ = 18.2±2.2pN (N=340). These values are in excellent agreement with previously reported results [48]. To quantitatively evaluate the unfolding, several force-ramp data sets were collected and analyzed using a model that converts the force rupture distribution to an intrinsic lifetime and distance to transition state [129] (Fig. 4-2B, 4-2C, and 4-2D). The model reveals that the distance to the transition state does not change appreciably in the presence of either osmolyte: $\Delta x^{\ddagger}_{Buffer}$ = 2.7±0.1nm, $\Delta x^{\ddagger}_{1M\ Sorbitol}$ = 2.8±0.1nm, and $\Delta x^{\ddagger}_{1M\ TMAO}$ = 2.3±0.1nm. Although the value for TMAO differs somewhat from the sorbitol and buffer values, this difference is likely an artifact of the model being sensitive to the shape of the distribution. Indeed, constant-force experiments (see below) reveal that $\Delta x^{\ddagger}$ is not significantly different under the three different conditions, consistent with previously reported findings [117]. Taken together, these results suggest that neither TMAO nor sorbitol change the unfolding pathway of T4*.

We note that the distance to the transitions state determined in our experiments, applying force to the termini of T4*, is unusually large. Most globular proteins exhibit distances to the transition state of less than 1 nm, reflecting the brittle nature of stably folded proteins [105]. Native T4* has a radius of gyration (R_g) of ~2nm [130]. The $\Delta x^{\ddagger}$ values determined here and previously [48] suggest that the molecule can be extended by approximately this length before crossing the barrier to unfolding. We believe that the origin of the large $\Delta x^{\ddagger}$ values is likely the unstable N-terminal A-helix region [131] that may deform easily under mechanical load before the barrier to unfolding is crossed (Fig. 4-1A). Regardless of the origin of the large absolute

values, $\Delta x^{\ddagger}$ appears to be the same in all cases, indicating that osmolytes do not appreciably affect the position of the barrier to unfolding.

The unfolding force distributions, analyzed as described above [129], suggest a change in the folded state lifetimes extrapolated to zero force (τ_o) when $\Delta x^{\ddagger}$ is fixed at 2.7nm. $\tau_{o,Buffer}$ = 46,101±188s, $\tau_{o,1M\ Sorbitol}$ = 69,916±261s, and $\tau_{o,1M\ TMAO}$ = 71,698±562s. Determining τ_o required extrapolation over a relatively large force range and the exclusion of transition state sliding, imparting some uncertainty onto the determined values. Nevertheless, given that osmolytes stabilize the native states of proteins, our results are not unexpected, as osmolytes have been shown to decrease unfolding rates both in bulk [110-112] and at the single-molecule level [114-117]. Taken together, the data are consistent with a slight decrease in the unfolding rate and no significant change in the pathway.

Similarly, osmolytes do not appear to impact the worm-like chain (WLC) model that is used to describe the unfolded state of proteins in force spectroscopy experiments [48, 97, 104]. We calculated the change in contour length (ΔL_c) for T4* unfolding in the presence and absence of TMAO and sorbitol. We did not detect any significant differences. The protein unfolds within a well-defined regime of extension changes and forces in the presence and absence of osmolytes (Fig. 4-2E). The calculated contour length changes in each case ($\Delta L_{c,Buffer}$ = 60.5±3.1nm, $\Delta L_{c,1M\ Sorbitol}$ = 59.9±3.2nm, and $\Delta L_{c,1M\ TMAO}$ = 59.6±2.6nm) are in good agreement with the expected value of 58.1 nm for T4* (i.e., 164 amino acids x 0.36nm per amino acid with a 0.9nm end-to-end distance in the crystal structure of the folded state gives 58.1nm). These results support the notion that the transitions observed in force-ramp experiments (lighter arrows, Fig. 4-2A) correspond to the cooperative unfolding of the entire molecule from a fully folded state. In addition we can rule out that osmolytes affect the persistence or contour length in our experiments, at least in the force range covered by the T4* unfolding transitions (~10 - 20 pN).

The data discussed above indicate that osmolytes do not affect the unfolding pathway, but may slow down unfolding rates. We directly measured unfolding rates in constant-force

experiments at different forces. An example trace is shown in Figure 4-3A. The molecule initially (t = 0) populates the native state and is held at 15pN until a cooperative unfolding transition occurs. Using data from multiple independent transitions, performed at different forces, we determined the distributions of lifetimes (τ_{app}) at each force, from which we calculated the force-dependent unfolding rate constants ($k_{app,unf}$) (Fig. 4-3B).

Over the experimentally accessible force range, we consistently observed a small deceleration of T4* unfolding rates in the presence of TMAO and sorbitol (Fig. 4-3B). Using Bell's model [132], we determined the force dependence of the rate constant and the distance to the transition state. The fits with and without osmolyte are not statistically different, yielding zero force unfolding rates (i.e. intercept) within error of each other and $\Delta x^{\ddagger}$ values (i.e. slope) that are only slightly larger than the values obtained from force rupture distributions: $\Delta x^{\ddagger}_{Buffer}$ = 3.24±0.17nm, $\Delta x^{\ddagger}_{1M\ Sorbitol}$ = 2.97±0.17nm, and $\Delta x^{\ddagger}_{1M\ TMAO}$ = 3.16±0.47nm. In short, the fits of the force dependence of the unfolding kinetics are within experimental error of one another, supporting the conclusions of the force ramp experiments, that the unfolding pathway is not affected by osmolytes.

Despite the similarity in the pathways, the unfolding rates at each individual force differ, with TMAO having the largest effect on the rate (Fig. 4-3B). The apparent unfolding rates at each force in the presence of Sorbitol and TMAO decrease 1.39±0.10 and 2.31±0.56 fold respectively (Tables 4-1 and 4-2), relative to buffer only. Because the apparent rate is related to the activation energy of the process, it is possible to determine the change in activation energy across the overlapped force regimes:

$$k_{app,unf} \propto \exp\left(\frac{-\Delta G^{\ddagger}}{k_B * T}\right) \qquad \text{(Eq. 4-8)}$$

The average difference in activation energy between these conditions is on the order of thermal noise: $\Delta\Delta G^{\ddagger}_{Sorb}$ = -0.33±0.07k_BT and $\Delta\Delta G^{\ddagger}_{TMAO}$ = -0.82±0.25k_BT. These changes in activation free energies and folding rates are smaller than what has been reported for other force

spectroscopy studies using AFM [115-117]. Importantly, our results indicate that at lower concentration (closer to a biologically relevant scenario) of osmolyte and at a lower force regime osmolytes do not affect unfolding kinetics as much as previously reported.

Osmolytes accelerate early events in folding

Previous force-spectroscopy experiments in the presence of osmolytes have been unable to directly observe folding because of the high force-regime required for AFM. We empirically noted that the refolding forces observed in the force ramp experiments (darker arrows in Fig. 4-2A) are slightly higher in the presence of osmolyte, suggesting faster folding rates. Unfortunately, reliable estimates of the refolding force could not be obtained in force-ramp experiments due to the noise in the measurement and hopping behavior at this low force regime. Thus, we carried out force-clamp experiments to directly observe folding and to accurately determine the folding rates of T4* (Fig. 4-3C). In these measurements, the molecule is populating the unfolded state at time zero and refolds after some time (denoted by the red arrow) while being held at a low, constant force. Performing similar analyses as was done for the unfolding transitions described above, we determined the force dependence of the apparent refolding rate constant $k_{app,fold}$ (Fig. 4-3D, Table 4-2).

Two major conclusions can be drawn. First, the apparent folding rate at a given force is accelerated the presence of osmolyte. The average fold increase in the presence of sorbitol and TMAO is 2.75±1.02 and 13.7±3.8 respectively. These values correspond to $\Delta\Delta G^{\ddagger}_{Sorb}$ = 0.96±0.36k_BT and $\Delta\Delta G^{\ddagger}_{TMAO}$ = 2.59±0.26k_BT, demonstrating a more pronounced effect than on the unfolding rate deceleration. Second, the folding pathway is not well described by a two-state transition, apparent from the non-linear dependence of $k_{app,fold}$ on force. This observation suggests the presence of a folding intermediate. Indeed, despite the increased noise at the lower forces required for folding, the presence of an intermediate state between the unfolded and the

native states is clearly apparent (blue arrow, Fig. 4-3C). The weaker force dependence at low forces (being due to the change in the average transition state) indicates that the transition into the intermediate state is more dominant at low forces. Conversely, a stronger force dependence is observed when transition into the native state is more dominant at high forces [105].

From the data it can be inferred that the intermediate observed in the single-molecule folding traces appears to be "on-pathway" and essentially irreversible once folded to the native state [48]. It should be noted that all the constant-force data are consistent with one intermediate state as previously reported [48]. Furthermore, upon addition of osmolyte the kinetics data are still consistent with one intermediate in addition to the binned distribution of distances revealing a single species. It is highly unlikely that multiple intermediates have the same kinetics, extension, and osmolyte sensitivity. This allows the reaction to be schematically represented as:

$$U \underset{k_{I\text{-}U}}{\overset{k_{U\text{-}I}}{\rightleftharpoons}} I \xrightarrow{k_{I\text{-}N}} N$$

where the unfolded (U) can exchange with the intermediate (I) until it irreversibly folds to the native state (N). To deconvolve the kinetic rate constants, $k_{U\text{-}I}$, $k_{I\text{-}U}$, and $k_{I\text{-}N}$ (where $k_{A\text{-}B}$ represents the rate of going from A to B), we implemented a Bayesian extension of a Hidden Markov Model (BHMM)[133]. Advantages of the BHMM are that it explicitly accounts for experimental noise in addition to distance fluctuations within each state, provides for state assignments as well as a deconvolution of the individual rate constants traces, and provides more precise values than HMMs by sampling the posterior (Fig. 4-4A).

At higher forces, $k_{I\text{-}U}$ increases, resulting in lifetimes of the intermediate that are on average too short for our experimental approach to detect. Lower forces, as well as the addition of osmolyte, increase $k_{I\text{-}N}$, similarly reducing the lifetime of the intermediate to the point that it cannot be reliably detected. Nonetheless, we were able to obtain reliable values for all three rate constants at 5.0pN (Fig. 4-4B). Due to the nature of the osmophobic effect (i.e. stabilizing folded compact states over more extended states), it may be expected that the rate of formation of either

the I state from the U state or the formation of the N state from the I state would be increased significantly as both the U to I and the I to N transitions presumably decrease exposure of the backbone to the solvent. Such a conclusion is not borne out of the current analysis. The increased apparent folding rate is instead due to a significant increase in $k_{U\text{-}I}$ (i.e., $k_{U\text{-}I,Buffer}$ =2.17 s^{-1} [1.88 s^{-1},2.29 s^{-1}] to $k_{U\text{-}I,1M\ Sorbitol}$ = 3.45 s^{-1} [3.00 s^{-1},3.93 s^{-1}] and $k_{U\text{-}I,1M\ TMAO}$ = 3.17 s^{-1} [2.57 s^{-1},5.30 s^{-1}]) and a decrease in $k_{I\text{-}U}$ (i.e., $k_{I\text{-}U,Buffer}$ =24.98 s^{-1} [22.03 s^{-1},26.00 s^{-1}] to $k_{I\text{-}U,1M\ Sorbitol}$ = 10.80 s^{-1} [9.51 s^{-1},14.08 s^{-1}] and $k_{I\text{-}U,1M\ TMAO}$ = 7.91 s^{-1} [6.25 s^{-1},12.50 s^{-1}]) (Fig. 4-4B).

Interestingly, the rate constant confidence intervals for the final irreversible step have significant overlap, revealing very little, if any, change in the intermediate to native state transition: $k_{I\text{-}N,\ Buffer}$ = 0.32s^{-1} [0.21s^{-1},0.48s^{-1}], $k_{I\text{-}N,1M\ Sorbitol}$ = 0.30s^{-1} [0.08s^{-1},0.65s^{-1}] and $k_{I\text{-}N,1M\ TMAO}$ = 0.86s^{-1} [0.32s^{-1},1.76s^{-1}]. Note, the values in brackets represent 95% confidence intervals as previously described[48, 133]. Although the increase in $k_{U\text{-}I}$ and decrease in $k_{I\text{-}U}$ does not explain the total change in the apparent kinetics (Fig. 4-3D), these two rates are clearly dominating the changes that are observed in the force-clamp experiments. The modulation of a specific rate constant in the single-molecule folding pathway of T4* has been reported previously. Specifically, the rate of N formation from I ($k_{I\text{-}N}$) is decreased when a nascent T4* version folds on the ribosome [48]. In direct contrast to the mechanism by which the ribosome affects T4* folding, osmolytes are promoting intermediate formation, an effect that increases the apparent folding rate significantly. Indeed, the population of the intermediate relative to the unfolded state increases significantly from 3.6% to 13.6% and 24.6% from buffer to sorbitol and TMAO, respectively (Fig. 4-5). Importantly, both classes of osmolytes affect folding by the same mechanism, albeit to different degrees (on a per molar basis).

Unique access to the intermediate state characterization

Since the single molecule force spectroscopy experiments conducted in the presence of osmolytes represent a two-dimensional chemo-mechanical perturbation of the protein, they provide a unique opportunity for characterizing the structural properties of the intermediate state. As noted above, analysis of the force-extension curves provides direct access to the contour length of the I and U states of T4* corresponding to 96-108 amino acids folded in the intermediate as previously reported [48]. We do not observe compaction of the intermediate or a deviation from the WLC model as a function of force for the intermediate supporting the assertion that the intermediate is the same state in the presence of osmolyte. However, we note that the probability of the intermediate changes significantly upon addition of osmolyte, increasing by almost an order of magnitude upon addition of TMAO (Fig. 4-5). As described previously [49-51], the effect of TMAO on an intermediate state is based on the amount of accessible surface area (ASA) relative to the unfolded state. Based on both the change in extension upon folding to the intermediate state, and the osmolyte dependence of the probability of this state, it should be possible to develop a detailed structural model of the intermediate species.

Structure-based calculations of ASA provide a means of assessing potential structural models of the intermediate based on which portions of the molecule are folded. Towards this end we analyzed the state distribution (i.e. probability) in the absence and presence of TMAO which reveals the change the equilibrium free energy of the intermediate relative to the unfolded state. Thus, the increase in the I-state population relative to the U-state with TMAO is explained by a Gibbs Energy of ~1.14kcal/mol (Fig. 4-5A and 4-5C). It should be noted that we could not find a reliable force regime in which U and I are in equilibrium without concomitant folding to the N-state under all conditions tested. To compare this free energy difference to potential structures, we developed two models, one being the unfolded state under mechanical force and the other being the intermediate state to compare their transfer free energies. The intermediate state was

treated as a contiguously folded portion of polypeptide flanked by unfolded regions. This is justified by the fact that the intermediate is on-pathway and hence native-like, thus using the native state structure for all subsequent calculations appears reasonable.

From such structural models, it is readily possible to calculate the ASA and to use known values to determine the Gibbs Energy of transfer to 1M TMAO for the species in consideration[49, 50]:

$$\Delta\Delta G = \Delta Gtr, U - \Delta Gtr, I = \sum_{i=AA\ type}[n_i \Delta g^o_{tr,i,SC}\, \Delta\alpha_i^{SC}] + \Delta g^o_{tr,BB} \sum_{i=AA\ type}[n_i \Delta\alpha_i^{BB}] \quad \text{(Eq. 4-9)}$$

Where $\Delta\Delta G$ is the transfer energy of the unfolded state relative to the intermediate, n_i is the total number of groups of amino acid (AA) type i, Δg^o_{tr} is the free energy of transfer for the side-chain (SC) or backbone (BB) to 1M TMAO, and $\Delta\alpha_i$ is the fractional change in solvent ASA from the unfolded to the intermediate state [49, 50].

To determine the $\Delta\alpha_i$ values in Eq. 4-9, estimates of the ASA of the unfolded and intermediate states are required:

$$\Delta\alpha_i^{bb\ or\ sc} = \frac{\sum_{j=1}^{n_i}(ASA_{i,j,U} - ASA_{i,j,I})}{n_i ASA_{i,Gly\cdot X\cdot Gly}} \quad \text{(Eq. 4-10)}$$

where the numerator is summed over all amino acids j of type i, $ASA_{i,jU}$ is the ASA of the unfolded state, $ASA_{i,j,I}$ is the ASA of the intermediate state, n_i is as defined in Eq. 4-9, and $ASA_{i,Gly.X.Gly}$ is the standard side-chain or backbone solvent asccessibility of Gly-X-Gly tripeptides presenting the maximally exposed surface area (X i the amino acid of type i) [125]. In the analysis, the folded portion of the intermediate states was assumed to adopt the same conformation as the crystal structure, and the unfolded segments were treated the same as the mechanically unfolded states described below.

The accessible surface area of mechanically unfolded T4* is presumably different from the chemically denatured state hindering structural modeling. The fully extended conformation can be readily calculated from physical chemistry principles, however, T4* is not fully elongated in the refolding experiments. At 5pN the unfolded state is ~25nm more extended than the native

state (Fig. 4-4A) as opposed to the fully elongated state of ~60nm. A simple linear interpolation between the native state and fully elongated state may provide a first order estimate of the ~25nm extended state ASA, but this is a radical simplification and assumes linearity. To address this concern, we performed steered molecular dynamics (SMD) to determine ASA and therefore the $\Delta\alpha_i^{bb\ or\ sc}$ dependence on extension between the native and fully elongated state. Briefly, the system was equilibrated and was subsequently pulled at a constant velocity until fully elongated, similar to previously published methods [134-136]. By performing multiple SMD simulations we were able to determine how $\Delta\alpha_i^{bb\ or\ sc}$ relates to extension and develop statistics on the ASA of the unfolded state (Fig. 4-6 and 4-7). From these values, we determined the free energy of transfer for the unfolded state ($\Delta G_{tr,U}$ in eq. 1) and that of the intermediate ($\Delta G_{tr,I}$ in eq. 1) as described above. Remarkably, the first order estimates of ASA from linear interpolation are close to the values calculated from SMD.

Using this method, we exhaustively enumerated all possible intermediate states for T4* and determined the free energy of transfer to 1M TMAO. These states were generated by systematically folding every possible contiguous portion of the molecule at the individual amino acid level, calculating the change in ASA upon folding, and then using the experimentally measured transfer free energies to obtain a Gibbs Energy. A plot representing all possible partially folded species with varying unfolded termini is shown in Fig. 4-8A. All species in the plot above the diagonal have the same free energy of transfer of 0kcal/mol because they are all fully unfolded. The intermediate candidates of interest in Fig. 4-8A are the ones in the bottom left half where there is a significant free energy difference. Several important features emerge. First there are clear differences in $\Delta\Delta G_{tr,U\text{-}I}$ for unfolding the N-terminus versus the C-terminus in 1M TMAO relative to the unfolded state (as evidenced by the asymmetry in Fig. 4-8A), which originates from differences in amino acid composition between the N and C terminal segments of a given length.

Second, when the energies are converted into the experimental changes in the population of the intermediate state, only a narrow portion of the potential unfolding combinations for the N and C termini agrees with experiment (green shaded region, Fig. 4-8A). Third, when the known change in contour length corresponding to 96-108 amino acids is included in the constraints (depicted as the yellow area between the two solid black lines in Fig. 4-8A), only one region is found where there is an overlap in both the experimental distance and population change constraints (red shaded region, Fig. 4-8A). The results suggest that the structural character of the intermediate state is most consistent with the C terminal domain being folded and the N terminal domain being unfolded (Fig. 4-8B, III). Remarkably, the predicted intermediate is consistent with two previous experimental observations. First, the division of the structure falls precisely at the known structural domain boundary of T4* previously identified by x-ray crystallography [137]. Second, the N-terminus being unfolded is consistent with the known relative stability of this domain obtained from both bulk experimental [138] and theoretical [131] studies.

We note that it has been previously suggested that the intermediate in the single-molecule experiments is the C-terminal domain [48]. However the results presented here reveal that a robust characterization is possible when single molecule experiments are performed in osmolyte solutions. The strength of this strategy lies in the fact that it can also be used to make quantitative predictions. For instance, Fig. 4-8B-III shows the intermediate that was determined from the analysis of the buffer and TMAO alone. Based on the known transfer free energies in sorbitol, the structural model can be used to predict the change in population of the intermediate in 1M sorbitol. Remarkably, the model predicts the population of the intermediate to within a tenth of a percent of the experimentally measured value (Fig. 4-8C). Importantly no other potential intermediates fall within the experimental error, providing further evidence of the robustness of the result, as well as the accuracy of the transfer free energy measurements used in this analysis [49-51]. Taken together, we have demonstrated that combining chemical perturbations with

single-molecule force spectroscopy provides a powerful tool for structural interpretation of force spectroscopy data that will likely be applicable to a wide variety of proteins.

4.5 Conclusions

This study represents the first direct observation of the osmophobic effect during protein folding at the single-molecule level using optical tweezers. We observe an increase in folding rate consistent with what has previously been reported for the effect of osmolyte. However, our ability to directly observe folding together with a robust method to determine individual rate constants from the single molecule traces (Fig. 4-4B) reveals surprising details of how osmolytes modulate T4* folding. Strikingly, formation of a folding intermediate is specifically modulated in the first step of the folding pathway: i.e. the population of the intermediate changes because k_{U-I} and k_{I-U} change, while k_{I-N} is unperturbed. This finding is unexpected because approximately one third of the molecule remains unfolded in the intermediate state, leaving a significant portion of the peptide backbone exposed to solvent. How is it possible, then, that k_{I-N} is not affected by osmolytes? Because the rate depends on how osmolyte is affecting the energy of the intermediate state relative to the transition state, the lack of a difference in k_{I-N}, with and without osmolyte, suggests that the surface area exposed in the transition state, relative to the intermediate state, is approximately the same.

Regardless of the origin of the effect, the functional implications are clear. Low concentrations of osmolytes decrease the rate of the transition from I to U, in effect minimizing the probability of an excursion to higher energy states. This suggests an osmolyte-driven dampening of dynamics of the protein, a result that is consistent with results obtained from hydrogen exchange studies [139]. We note that our results report the effect of osmolyte on one protein (T4*). The fact that we obtain similar results using two different osmolytes applied to one protein of course does not mean that the results will be the same for all proteins.

Nonetheless, the results presented here reveal that single molecule force-spectroscopy provides a powerful tool to understand the impact of osmolytes on protein stability.

We note that at least in the case of T4*, osmolytes primarily affect the rate of the U to I transition, while the ribosome was recently shown to affect the I to N transition [48], indicating that osmolytes and ribosomes exert their impact on different parts of the T4* folding reaction. Whether this effect is a general phenomenon, remains to be seen. However, it is interesting to speculate on the benefits of biological systems being able to evolve orthogonal mechanisms that presumably can be effective both together and separately. Did osmolytes evolve so as to minimize the potential of adversely affecting co-translational folding in organisms where osmolytes are co-localized with functioning ribosomes?

Perhaps most importantly, our results demonstrate that using osmolytes as a chemo-mechanical perturbation is a powerful approach for gleaning structural information about intermediates in the form of a second metric (i.e., solvent accessible surface-area). Although applied here to data obtained from optical tweezers, the strategy is not limited to force spectroscopy. Other techniques, such single-molecule fluorescence spectroscopy, providing access to changes in populations of states in different osmolyte concentrations, can, in principle, also provide information about surface area. Whether this is indeed the case awaits further studies.

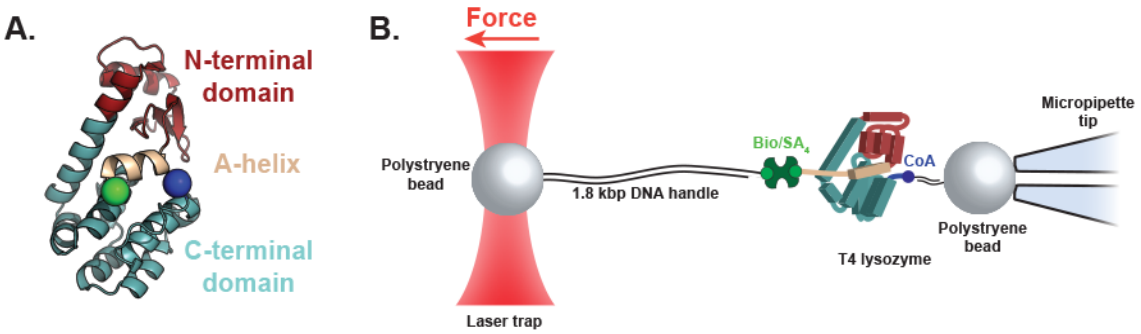

Figure 4-1. T4 Lysozyme as a model system and single molecule folding experimental setup A. Cartoon representation of T4 Lysozyme (PDB ID: 2LZM). Highlighted is the N-terminal A-helix (beige, residues 1-11), the N-terminal subdomain (red, residues 12-66), and the C-terminal sub-domain (cyan, resides 67-164). Also shown are the attachment points for single-molecule force spectroscopy experiments (green and blue spheres). **B.** In the experimental setup a polystyrene bead is held in the trap that exerts force on the single molecule. The bead is covalently linked to 1.8kbp of dsDNA which is attached to biotin (green circle) at the other end. There is a non-covalent biotin/streptatvidin interaction that then is linked to T4. CoA (blue circle) is covalently linked to T4 and another 50bp of dsDNA that is attached to the polystyrene bead which is held by suction.

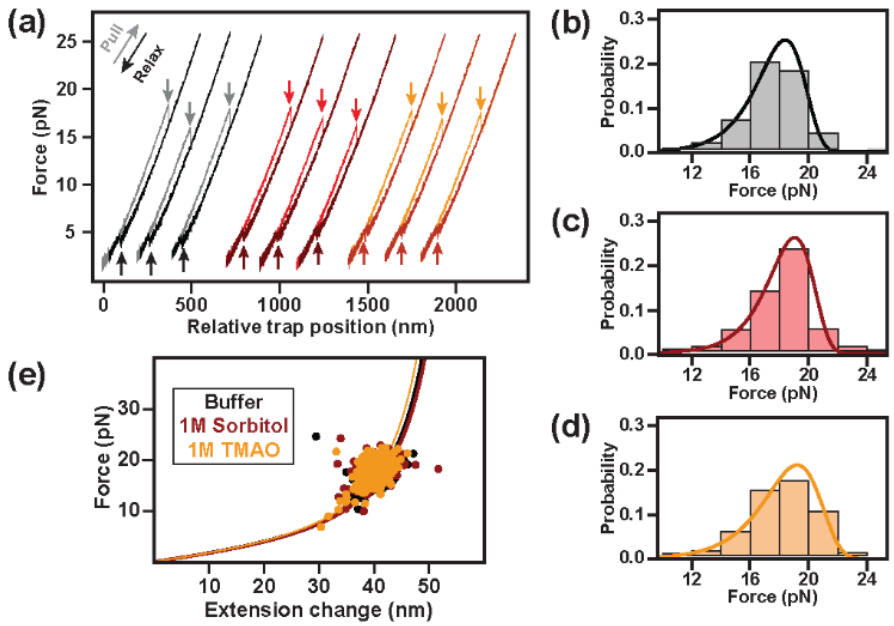

Figure 4-2. The mechanical unfolding pathway is unaffected by osmolyte A. Shown are representative force-extension curves (FECs) for single molecules of T4 lysozyme generated under constant velocity (100nm/s). FECs in buffer (black), sorbitol (red), and TMAO (orange) behave similarly. Unfolding transitions (lighter arrows pointing down) were observed in the 12-24pN regime for all conditions. Refolding occurred in the 3-6pN regime (darker arrows pointing up). B, C, and D show force rupture probability distributions in the presence of buffer (black), sorbitol (red), and TMAO (orange) respectively. Overlaid on the distributions are the fits using a theoretical model that yields distances to the transition state and lifetimes of the folded state [129]. E. Shown are the unfolding transition extension changes (nm) versus the forces of unfolding (pN) for every transition observed. Data are color coded similar to (A) and are fit to the wormlike-chain model [92]. All contour length changes (ΔL_C) that are consistent with unfolding of the whole molecule and are within error of each other.

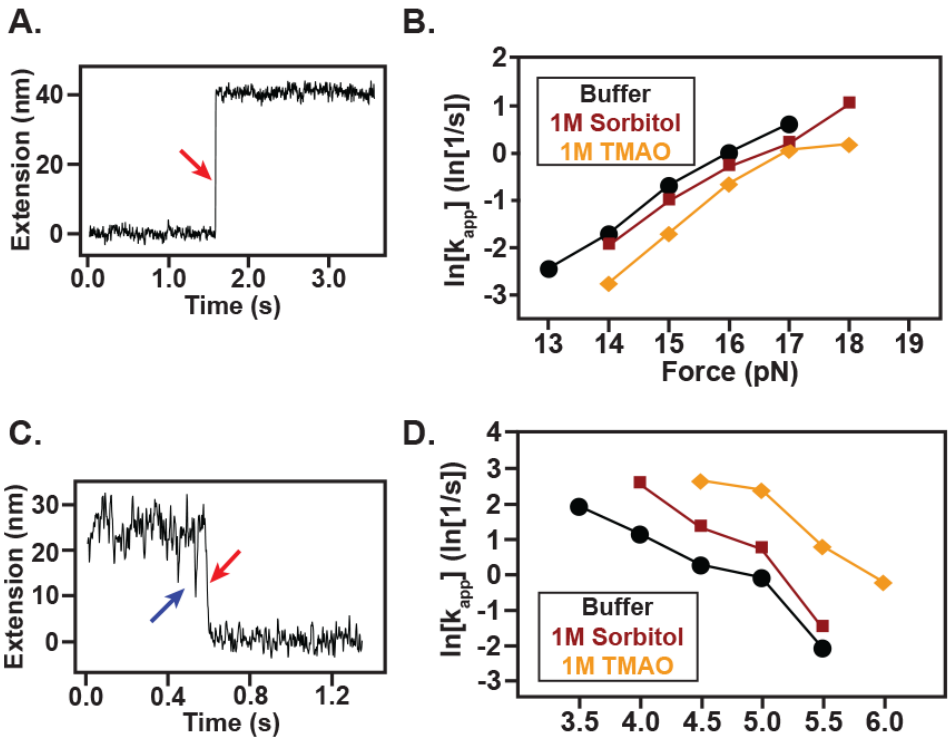

Figure 4-3. Osmolytes affect protein folding kinetics A. Representative force-clamp unfolding transition. Shown is the extension as a function of time with the unfolding transition denoted by the red arrow. Unfolding results in an increase in the relative position of the trap. **B.** Apparent unfolding kinetics as a function of force. Shown are the apparent kinetic rates under all three conditions tested. Error bars are smaller than the plot points and are thus not visible. **C.** Representative force clamp folding transition. Shown is the extension as a function of time with the folding transition denoted by the red arrow. There is a statistically significant on-pathway intermediate observed which is denoted by the blue arrow. **D.** Apparent folding kinetics as a function of force. Shown are the apparent kinetic rates in all three conditions tested. Due to the significant change in the rates it was not possible to obtain complete overlap of the force regime. Error bars are smaller than most of the points on the plot and are thus not visible.

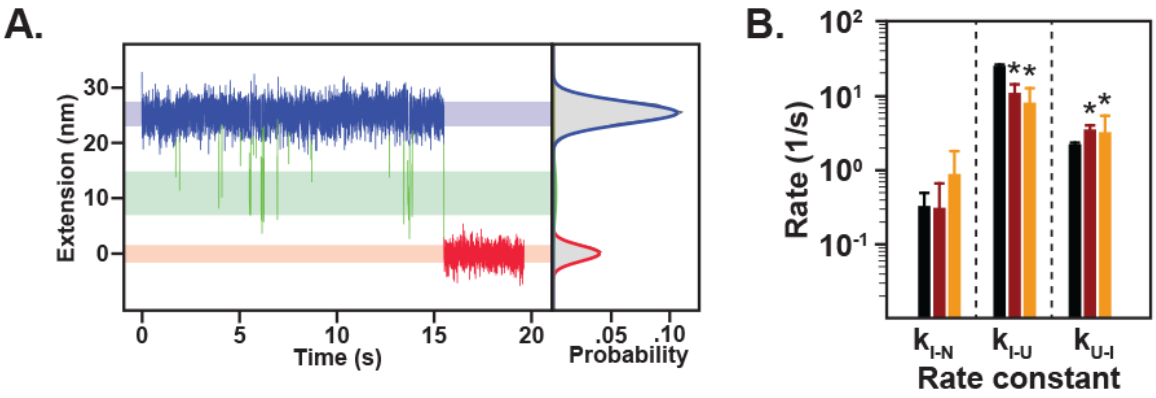

Figure 4-4. Osmolytes specifically affect the first step in folding. A. Representative refolding trace (force clamp at 5pN) fit to the BHMM. The graph shows the extension data with states assigned according to the BHMM: the unfolded state is shown in blue, the intermediate in green, and the native in red. The width of the shaded bars under the data represents the mean plus or minus one standard deviation which in this case is: $\mu_U \pm \sigma_U$ = 25.3±2.0nm, $\mu_I \pm \sigma_I$ = 10.9±3.8nm, $\mu_N \pm \sigma_N$ = 0.0±1.4nm. On the right is the probability distribution of the states in this refolding transition. **B.** Fits of each rate in the folding transition as determined by the BHMM for buffer (black), sorbitol (red), and TMAO (orange). The rates showing significant change are the k_{I-U} and k_{U-I} rates denoted by the asterisks – $p < .05$.

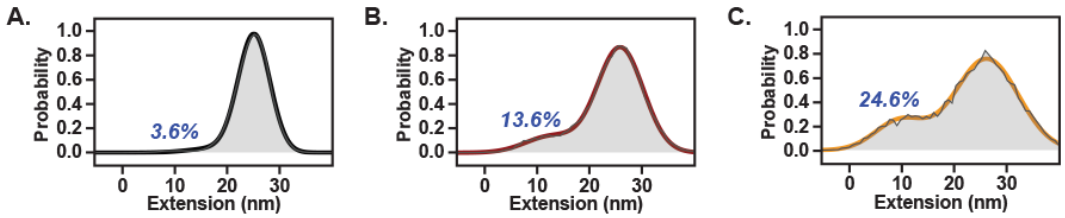

Figure 4-5. Osmolytes stabilize formation of the intermediate. A, **B**, and **C** are binned extension distributions for all transitions used for fitting from buffer (black), sorbitol (red), and TMAO (orange) respectively. The mean values of extension have not changed appreciably, however the relative populations of both have changed significantly from 3.6% to 13.6% and 24.6% from buffer, to sorbitol and TMAO respectively.

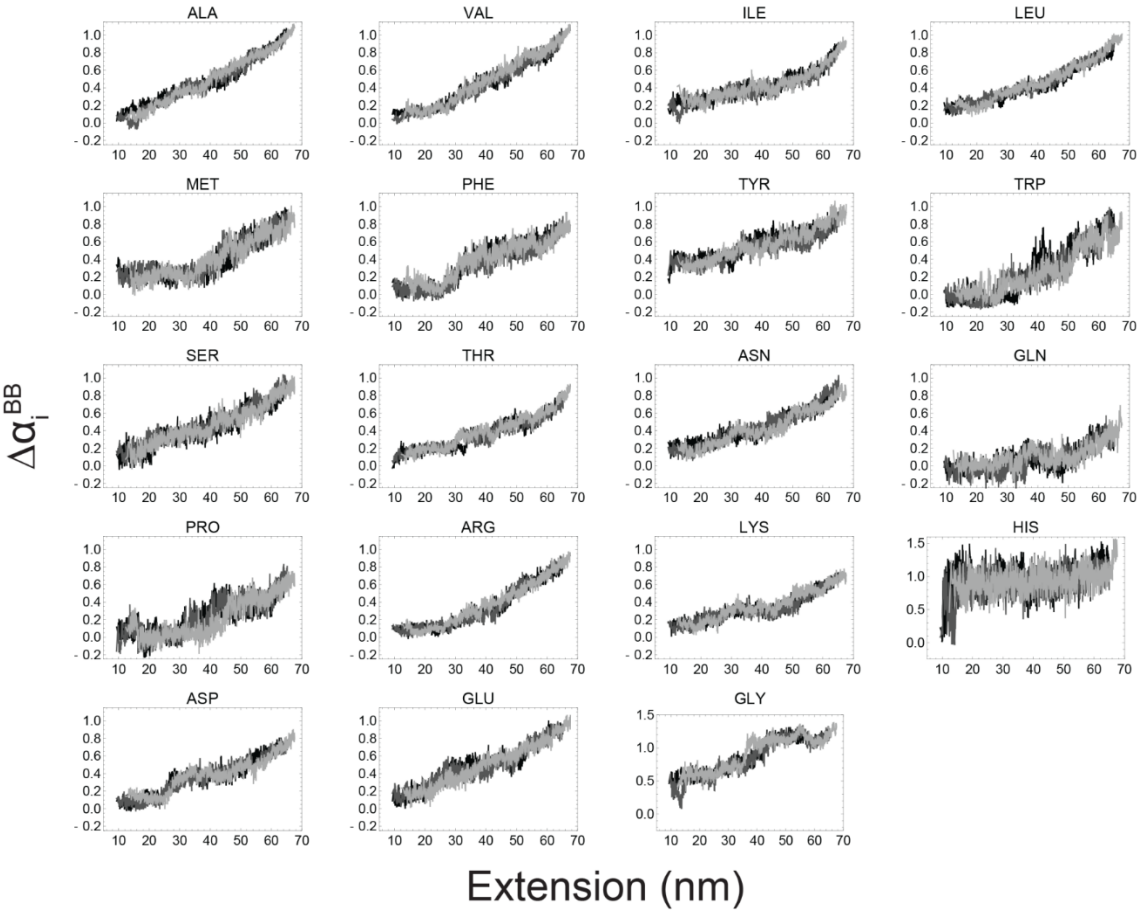

Figure 4-6. Backbone exposed surface area upon mechanical unfolding. Shown are the calculated Δα values for all amino acids in T4* from the steered molecular dynamics simulations. These values are for the backbone (BB) and all three simulations are overlaid showing excellent reproducibility. Note that the majority of amino acids are simply a line between the initial Δα from the crystal structure to 1.0 when fully extended. Average values corresponding to the dimensions of the unfolded state from the BHMM were used in transfer free energy calculations of the intermediate.

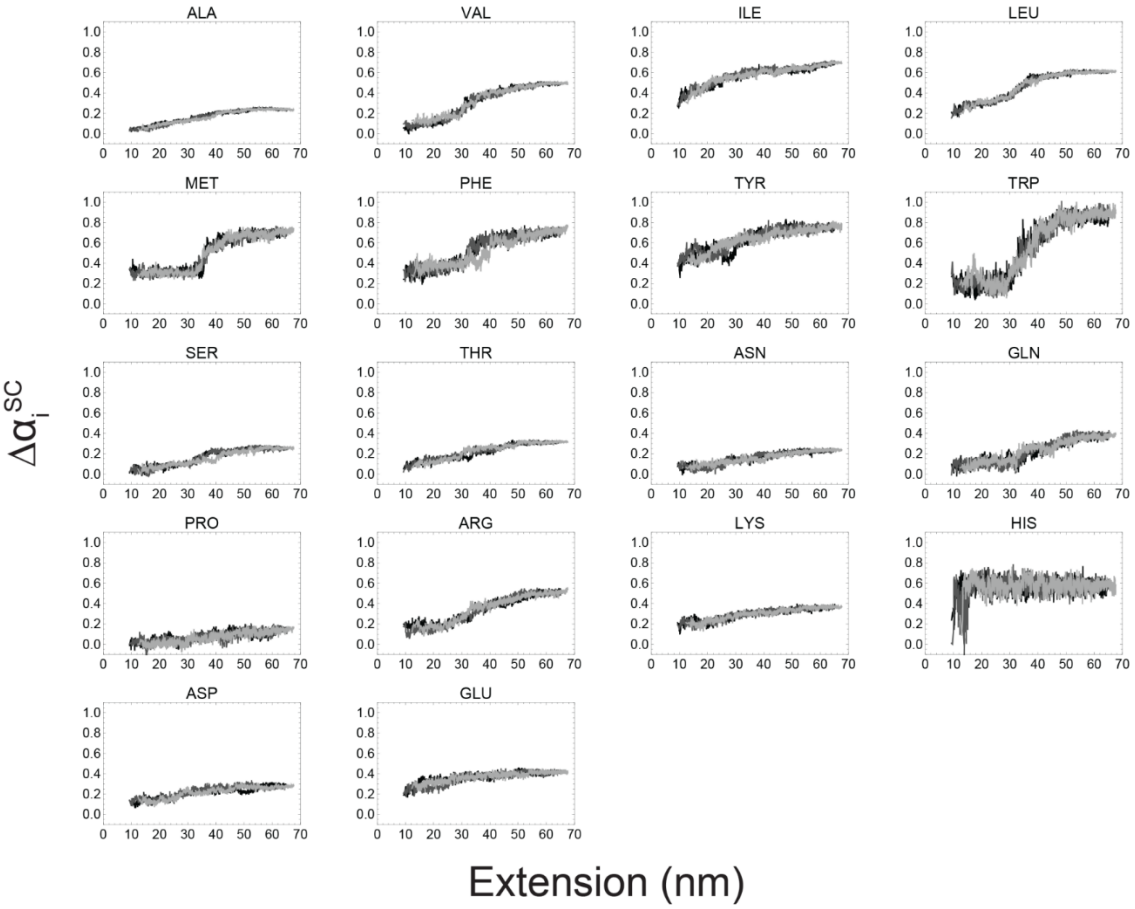

Figure 4-7. Side-chain exposed surface area upon mechanical unfolding. Shown are the calculated Δα values for all amino acids in T4* from the steered molecular dynamics simulations. These values are for the side chains (SC) and all three simulations are overlaid showing excellent reproducibility. Note that the majority of amino acids are simply a line between the initial Δα from the crystal structure to the maximum accessibility when fully extended. Average values corresponding to the dimensions of the unfolded state from the BHMM were used in transfer free energy calculations of the intermediate.

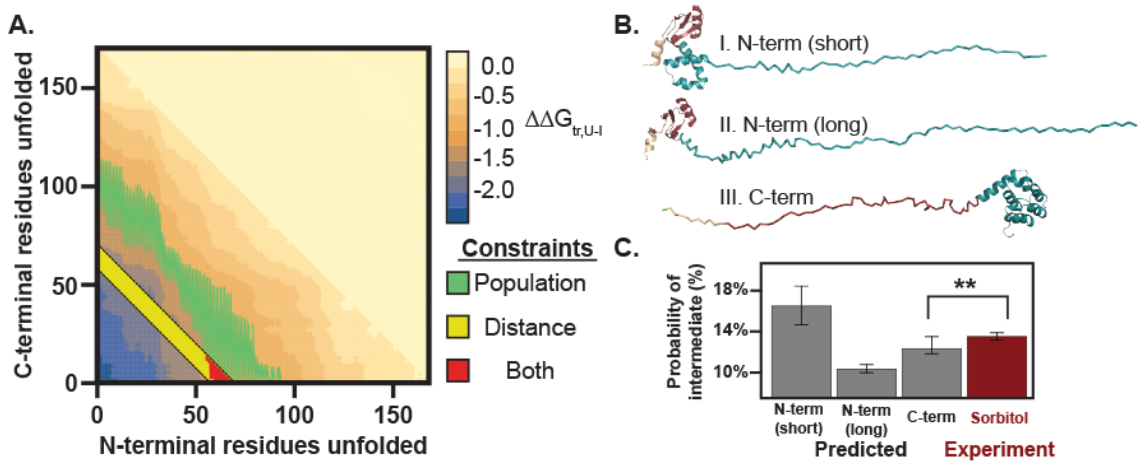

Figure 4-8 – Structural model of the intermediate and its predictive power. A. Shown is the plot of $\Delta\Delta G_{tr,U\text{-}I}$ for every possible intermediate of T4*. The axes represent the portion of the N-terminal or C-terminal residues unfolded in the intermediate. Shaded in green are intermediates that are consistent with the population in 1M TMAO ($\Delta\Delta G_{tr,U\text{-}I}$ = -1.14±0.20kcal/mol). The yellow region is the dimensional restraints from the BHMM. Shaded in red are the intermediates consistent with both restraints. These are all intermediates where the N-terminal domain is mostly unfolded and the C-terminal domain is primarily folded. **B.** Shown are potential structural models of intermediates. Color coding is identical to Fig. 1A. **B-I**. Depicted is the N-terminal subdomain that is consistent with the population constraint (folded amino acids 60-72, denoted N-term short). **II.** Depicted is the N-terminal subdomain portion along with C-terminal subdomain portion folded to be consistent with distance restraint (folded amino acids 96-108, denoted N-term long). **III.** Depicted is the C-terminal subdomain structural model that is consistent with both constraints (C-term). **C.** Quantitative prediction of the probability of the intermediate. The mean and ranges of populations consistent with

these domain boundaries are 16.5% [14.7%,18.5%], 10.4% [10.0%,10.8%], and 12.4% [11.9%,13.5%] for N-term long, N-term short, and C-term respectively. The final bar is the measured population of the intermediate in 1M sorbitol (red): 13.6% [13.2%,13.9%]. The only intermediate consistent with the experimentally measured value is the proposed C-terminal domain denoted **.

Table 4-1 – Unfolding kinetics upon addition of osmolyte

Condition	HKM		Sorbitol				TMAO			
Force (pN)	k_{app} [a]	N [b]	k_{app} [a]	N [b]	Change [c]	ΔΔG(kb*T) [d]	k_{app} [a]	N [b]	Change [c]	ΔΔG(kb*T) [d]
13	0.0854 ± 0.0008	67	N/A				N/A			
14	0.1781 ± 0.0025	100	0.1383 ± 0.0035	59	1.29	-0.25	0.0628 ± 0.0006	124	2.84	-1.04
15	0.4945 ± 0.0041	140	0.3527 ± 0.0067	73	1.40	-0.34	0.1796 ± 0.0010	186	2.75	-1.01
16	0.9898 ± 0.0068	138	0.7290 ± 0.0071	81	1.36	-0.31	0.5123 ± 0.0033	132	1.93	-0.66
17	1.8131 ± 0.0125	157	1.1872 ± 0.0132	79	1.53	-0.42	1.0446 ± 0.0150	131	1.74	-0.55
18	N/A		2.7291	0.0965	21	N/A	1.1807 ± 0.0192	74		N/A
			Average		1.39	-0.33		Average	2.31	-0.82
			StdDev		0.10	0.07		StdDev	0.56	0.25

Table 4-2 – Folding kinetics upon addition of osmolyte

Condition	HKM		Sorbitol				TMAO			
Force (pN)	k_{app} [a]	N	k_{app} [a]	N [b]	Change [c]	ΔΔG(kb*T) [d]	k_{app} [a]	N [b]	Change [c]	ΔΔG(kb*T) [d]
3.5	6.6049 ± 0.1382	29	N/A				N/A			
4	3.0043 ± 0.0316	92	12.3369 ± 0.3832	59	4.11	1.41	N/A			
4.5	1.2467 ± 0.0263	154	3.5842 ± 0.0848	55	2.87	1.06	13.6907 ± 0.1844	106	10.98	2.40
5	0.8686 ± 0.0060	236	1.9445 ± 0.0533	112	2.24	0.81	10.4996 ± 0.1715	264	12.09	2.49
5.5	0.1177 ± 0.0062	60	0.2073 ± 0.0118	66	1.76	0.57	2.1226 ± 0.0468	84	18.04	2.89
6	N/A		N/A				0.7654 ± 0.0091	136		N/A
			Average		2.75	0.96		Average	13.70	2.59
			StdDev		1.02	0.36		StdDev	3.80	0.26

[a]The reported values for k_{app} are the apparent rates that describe the single-exponential lifetime distributions at each force for unfolding (Table 4-1) and folding (Table 4-2).

[b]N is the number of transitions observed at that force across all molecules. The approximate number of transitions from each molecule was approximately the same and thus the statistics are robustly determined.

[c]The reported change in the rate constant is calculated relative to HKM (Buffer) conditions.

[d]The change in activation free energy is calculated based on the relative change of the apparent rate constants.

Chapter 5 – Single-molecule force spectroscopy on disordered proteins in the presence of osmolytes

Portions of this work have been published as:

Motlagh HN, Anderson JA, Li J, and Hilser VJ*. (2015) Disordered allostery: lessons from glucocorticoid receptor. Biophysical Reviews, **104** (2) 402a.

Motlagh HN, Li J, Thompson EB, and Hilser VJ*. (2012) Interplay between allostery and intrinsic disorder in an ensemble. Biochemical Society Transactions, **40** (5) 975-80.

Li J, **Motlagh HN**, Chakuroff C, Thompson EB, and Hilser VJ*. (2012) Thermodynamic dissection of the intrinsically disordered N-terminal domain of human glucocorticoid receptor. Journal of Biological Chemistry, **287** (32) 26777-87.

*To whom correspondence should be addressed: hilser@jhu.edu

5.1 Abstract

The ensemble basis of allostery has been known for years, yet only recently has emphasis shifted from interpreting allosteric mechanism in terms of discrete structural pathways to ones that focus on the statistical nature of the signal propagation process. This provides a vehicle to unify allostery in structured, dynamic, and disordered systems. In particular, intrinsically disordered (ID) proteins (IDPs), which lack unique stable structure, have been directly demonstrated to exhibit allostery in numerous systems, a reality that challenges traditional structure based models that focus on allosteric pathways. In this chapter we focus on studies from human Glucocorticoid Receptor (GR), a member of the steroid hormone receptor (SHR) family. The numerous translational isoforms of the disordered N-terminal domain of GR consist of coupled thermodynamic domains that contribute to the delicate balance of states in the ensemble, and hence *in vivo* activity. We find that these isoforms fold cooperatively upon the addition of osmolyte to a set of states that are thermodynamically indistinguishable from globular proteins. We probe the folded energy landscape of GR by performing single-molecule force spectroscopy experiments on various isoforms. The results reveal that IDPs are amenable to single-molecule force spectroscopy. The data are rich in information, but are too complex to be state assigned reliably. We discuss general conclusions and future directions for dissecting the folded landscape while incorporating these results. The ensemble nature of allostery illuminated from these studies lends credence to the models proposed and provides ground rules for allostery in all systems.

5.2 Introduction

The structure-function paradigm has been challenged by the increasing number of proteins found to contain intrinsically disordered (ID) regions [140-143]. Many of these ID regions undergo coupled folding and binding which leads to the presumed biologically active state(s) [140, 143-146]. What is also clear is that there are many ID sub-regions that are coupled to one another within the context of a larger ID polypeptide [54, 147], yet how these coupling energies are tied to function is largely unknown. Dysregulation of these ID regions leads to a variety of diseases, thus studying them at a basic level will glean insight into the molecular basis of such disease states [144, 148, 149]. Furthermore, the prevalence of IDPs and IDRs in transcription factors suggests that disorder can facilitate allosteric coupling. Unfortunately, experimental data supporting this hypothesis has been sparse until recently [43, 150, 151]. Studying the folded conformation for most ID proteins (IDPs) presents a technical challenge as well because not all cofactors are known and there may be conditional dependence for proper folding. How do we study the folded biologically active conformation(s) of IDPs in the absence of their binding partners? How do we probe the role of an ensemble and its coupling to allostery in disorder?

The SHR family of transcription factors is an ideal set of systems to study the interplay between allostery, intrinsic disorder, and function. SHRs consist of a conserved domain architecture consisting of an ID N-terminal domain (NTD), a DNA binding domain (DBD), and a C-terminal ligand-binding domain (LBD) (Fig. 5-1A) [152]. Strikingly, there is no sequence conservation between the NTDs of the SHR family yet they all contain an activation function (AF) region required for full transcriptional activity [153]. In GR the first AF region (AF-1) is found in the disordered NTD. For the human GR, eight different translational isoforms have been reported that vary in length, tissue distribution, and *in vivo* activity (Fig. 5-1B) [154].

Interestingly, the start sites of some isoforms of the NTD of GR are outside of the AF-1 region suggesting biological pressure to maintain these sites (Fig. 5-1B).

A perplexing question has been what functional roles do the conserved start sites outside of the AF-1 region serve? Previous work has revealed that coupling exists between different ID regions outside of the AF-1; specifically the N-terminal and C-terminal regions flanking the AF-1 are negatively coupled as determined by *in vivo* activity [155]. Interestingly, it was even shown that the ID-NTD is also coupled to the structured DBD such that structure is gained in the AF-1 that preferentially binds certain co-regulatory binding partners [156]. These results qualitatively indicate that there is coupling and suggest that gaining structure in the ID-NTD is required for biological function, yet a quantitative measurement of these couplings and a predictive model for folding remain elusive.

This chapter touches on previous measurements on the thermodynamic coupling between domains contained in the ID NTD of GR by discussing previously published bulk refolding experiments with the protective osmolyte TMAO. Our results suggest that the refolded conformation is thermodynamically similar to globular proteins based on structure-based calculations. We hypothesized that this would make IDPs amenable to single-molecule force spectroscopy. Indeed, this is the case and what appears to be a two-state transition in bulk is multi-state at the single-molecule level. It is apparent that the behavior of GR's NTD under force is extremely complex which hinders reliable state assignment on the constructs tested in this chapter. We discuss general conclusions and future directions for dissecting the mechanical landscape further.

5.3 Materials and Methods

TMAO induced protein folding transitions

These experiments were performed by another member of the lab and all methods are detailed as previously described [54]. These methods were used to refold the single-molecule constructs *in situ*, i.e. in the microfluidics chamber.

Assessment of side chain and backbone contributions to the m-value for TMAO-induced folding

M-values were calculated based on the group transfer free energy (GTFE) model [157] using experimentally measured values for the backbone and the side chain contributions [158, 159]. Briefly, the m-value considers each residues contribution using the difference in exposed surface area (SA) to a 1.4Å rolling sphere in the native state (crystal structure) relative to the denatured state (average SA from library of Gly-X-Gly rotamers):

$$m_{calc} = \Delta G_{tr,D} - \Delta G_{tr,N} = \Sigma_{i=AA\ type}[n_i\, \Delta g^{o}_{tr,i,SC} \Delta\alpha^{SC}_i] + \Delta g^{o}_{tr,BB}\, \Sigma_{i=AA\ type}\, [n_i \Delta\alpha^{BB}_i] \qquad \text{(Eq. 5-1)}$$

where n_i is the total number of groups of type i present in the protein, Δg^{o}_{tr} is the free energy of transfer of the side chain (SC) or backbone (BB) from aqueous solution to 1M TMAO, and $\Delta\alpha_i$ is the fractional change in solvent accessibility from native to denatured [158, 159]. Crystal structures with diverse folds chosen for analysis were RNase T1, P Protein, *Staphylococcal* nuclease, spermwhale myoglobin, carbonic anhydrase, adenylate kinase, triose phosphate isomerase, prophobilinogen deaminase, map kinase, phosphoglycerate dehydrogenase, and seryl tRNA synthetase (PDB IDs in Table 5-1). Only the A-chain was considered for calculations in the asymmetric unit and any bound metals and/or ligands were removed. For sensitivity analyses, every amino acid (i.e. there is 0% sequence identity between mutant and WT sequences) was changed based on expected probabilities from the BLOSUM62 matrix [160] and a similar

substitution matrix for order to disorder mutations [161]. 100,000 *in silico* mutant sequences for each protein were generated and the m-value were recalculated using the same fractional change in solvent accessibility (Table 5-1).

Protein purification for single-molecule experiments

A single molecule construct of GR isoform A-NTD was ordered from DNA 2.0 (Menlo Park, CA) such that it was flanked by an Avi Tag on the N-terminus and an YbbR tag on the C-terminus. The construct also contained ten tandem histidines followed by a TEV protease site on the N-terminus for purification. This construct was used as a template to subclone all other isoforms of the NTD by means of Gibson Assembly [162, 163]. All constructs were expressed and purified identically as described below.

Protein was expressed by transforming into Rosetta™ 2(DE3)pLysS (EMD Millipore) cells, outgrowing to preparative quantities until OD_{600} = ~0.6-0.8, and induced overnight at 18°C with 1mM IPTG. Cells were pelleted after 18 hours of induction and frozen at -80°C. The cell pellets were resuspended in denaturing lysis buffer (6M guanidine hydrochloride, 100mM monosodium phosphate, 10mM Tris, 20mM imidazole, pH 8.0). Cells were lysed by sonication and clarified by centrifugation. The supernatant was then loaded onto a nickel agarose column, renatured directly on the column (same composition as lysis buffer except it does not contain guanidine hydrochloride), and eluted with an imidazole gradient. The protein was then TEV digested and purified in a manner identical to that described in the previous chapter for T4 Lysozyme. The protein was finally dialyzed to storage buffer (10mM HEPES, 80mM NaCl, 1mM EDTA and 10% glycerol, pH 7.6), concentrated, flash frozen with liquid nitrogen, and stored at -80°C. Protein was labeled for single-molecule experiments as described in the previous chapter.

5.4 Results and Discussion

Intrinsically disordered proteins are thermodynamically similar to globular proteins

It is well established that certain IDPs undergo coupled folding and binding and that the folded state is the functional conformation [26, 164]. Indeed, cooperative folding of the AF-1 of GR with the osmolyte trimethylamine *N*-oxide (TMAO) increases its affinity for known binding partners as well as protection from protease digestion suggesting that the TMAO refolded state is the functional conformation that is well folded [165]. Data collected by another member of the lab demonstrate that the A, C2, and C3 isoforms of the GR NTD also fold cooperatively in the presence of TMAO, a hallmark of a spontaneously folding protein (Fig. 5-2A) [150]. What is clear from analysis of Figure 5-2A is that the three isoforms in question have different stabilities and that removal of the residues between A and C3 increases the stability of the C3 isoform (Fig. 5-2B). Moreover, when the live cell culture activities of these isoforms are plotted against the measured stabilities there is a clear relationship (Fig. 5-2B). Taken together, these results support three hypotheses: (1) that the GR NTD undergoes coupled folding and binding to perform its biological role, (2) that a contributing factor to this process is the energetic cost of folding, and (3) the different isoforms can modulate the stability and hence biological activity of this process [150]. This led us to propose the model that the NTD consists of at least two thermodynamically coupled domains: a regulatory (R) region and a functional (F) region (Fig. 5-1) [150].

Closer analysis of the TMAO induced refolding data reveals that there is a relationship between the m-value and the length of the protein (Fig. 5-2C). The m-value describes the stability of the protein and its dependence on the concentration of denaturant or osmolyte [166]:

$$\Delta G_{Protein}(x) = \Delta G_{H20} - m * [x], \qquad \text{(Eq. 5-2)}$$

where $\Delta G_{protein}$ is the stability at a given concentration of osmolyte/denaturant, ΔG_{H20} is the stability of the protein in the absence of osmolyte/denaturant, $[x]$ is the concentration of

osmolyte/denaturant, and m is the m-value as described. What can be appreciates it that the GR-NTD isoforms fall on the same line as globular proteins. Given that the m-value from TMAO corresponds primarily to backbone burial [49], this result suggests that the refolded conformation of the GR-NTD isoforms is thermodynamically similar to a globular protein fold. In other words, when GR folds, it buries as much backbone surface (on a per residue basis) as globular proteins. The fact that the full length and different length truncated isoforms fall on the same line, suggests that not only the F-region (C3-NTD) folds to a globular structure, but that the R-region (included in the A-NTD) does as well (Fig. 5-2C, gray circles).

We note that TMAO-induced m-values appear to be primarily determined by energetically unfavorable interactions between TMAO and the peptide backbone, and are expected to be affected to a lesser degree by side chain contributions [158]. This expectation is especially important given that ID proteins have more charged residues, less hydrophobic residues, and lower sequence complexity compared with globular proteins [142, 143]. To test whether the sequence differences between ID and globular proteins would dramatically impact the measured m-value, and thus the interpretation of the similarity in m-values presented here, m-values for a database of globular proteins were computed using artificial sequences generated in two ways. First, sequences were generated using property-based similarity (using BLOSUM62 [160]). This produces sequences with similar properties to the parent structure, but with the amino acid usage characteristic of globular proteins. Second, sequences were generated by randomly assigning amino acids to positions keeping the distributions consistent with ID proteins [161]. As Figure 5-3 reveals, regardless of the model chosen, the amino acid composition does not dramatically impact the expected m-value.

The importance of the results presented in Figure 5-3 is two-fold. First, they reveal that sequence differences between globular and ID proteins are not a strong determinant of the measured m-values, and thus validates the apparent similarity between the folded states of

globular and ID proteins inferred from m-values. Specifically, ~3/4 of the m-value contributions arise from the backbone while the remaining 1/4 stems from the specific side chains (Table 5-1). Second, and perhaps most important, they provide an important new tool to quantitatively compare the folded or native state(s) of both ID and globular proteins. To date, attempts to study differences between globular and ID proteins have focused on the unfolded states [167], a bias that stems from the uncertainty as to whether all ID proteins need to fold in order to function, and the difficulty in identifying conditions that can facilitate folding for those ID proteins. Until now, such factors have undermined the idea that ID proteins have the capacity to adopt folded structures similar to their globular counterparts, or that their thermodynamics could even be quantitatively compared.

Intrinsically disordered proteins are amenable to single-molecule force spectroscopy

The results from the previous section suggest that there are multiple coupled domains contained within the ID-NTD of GR and that they are globular-like when refolded by osmolyte. It is clear that that the coupling, energetic stability, and rates of folding are contributing factors to *in vivo* activity. How can we make measurements that report on these properties? An issue with IDPs in general, and GR's NTD specifically, is that they are aggregation prone [45]. Indeed, the experiments conducted in the previous sections in bulk were at 0.5μM. The protein cannot be concentrated to a regime required for most solution biophysics experiments [44]. This rules out any structural characterization via NMR or crystallography and severely limits the bulk biochemical techniques to our disposal. One way to circumvent this issue is to perform single-molecule experiments.

Since GR's NTD and its isoforms are thermodynamically similar to globular proteins when they are folded, we would expect them to be amenable to single-molecule force-

spectroscopy. This suggests that we would be able to directly observe the ensemble nature of IDPs at the single molecule level. This would help us deconvolve the intermediates we expect to observe and would yield access to folding rates and the coupling between different IDRs if we pull on the various isoforms.

To test this hypothesis, we pulled on various isoforms of GR's NTD in the presence and absence of osmolyte in a manner similar to those described in the previous chapter (Fig. 5-4). Since intrinsically disordered proteins lack a stable tertiary structure, it would be expected that mechanically pulling on GR's NTD would result in no structural transitions. Indeed, when pulling on the A-NTD we reproducibly observe no significant structural transitions (Fig. 5-5B). This result is also observed for the C3-NTD and D1-NTD (Fig. 5-5). This result is also reproducible across a large number of molecules (Table 5-2 – Row 2).

Despite the lack of transitions, a fair question is: how do we know that these tethers contain protein? It is not uncommon to generate "empty tethers" – i.e. only with the 1.8kbp DNA handle and without protein attached to the 50bp DNA spacer. Despite numerous efforts we have been unable to identify how this is mechanistically viable, especially since the experimental setup suggests this should not be possible. We found that it has been able to discern between empty tethers and those containing protein by three independent pieces of information. The first is that the presence of an unfolded protein changes the curvature of the entropic elasticity [92]. This is a small signal difference and must be analyzed by hand after the experiment.

More telling is the lifetime of the tether once DNA overstretching at 65pN has been reached. It is now well-established that once overstretching occurs, the DNA handles no longer have stacking interactions because they are in a "ladder" like conformation [92], and as such, are only held together by the hydrogen bonding between base pairs. As a result, the rate of dissociation at the overstretching force is proportional to the number of hydrogen bonds, and thus,

the number of base pairs. If an empty tether is generated, the DNA handle is ~1.8kbp long and can be pulled through overstretching completely. However, if it contains protein, the 50bp of DNA on the C-terminus will dissociate almost immediately and cannot go through overstretching. This does not rule out that there would be some overlap between the dissociation rate kinetics, but the odds of this occurring reproducibly are negligible.

Perhaps the most compelling result that indicates we are obtaining tethers with protein occurs when we add osmolyte to a concentration at which we expect the molecule to be folded. If the molecule is cooperatively folding upon addition of osmolyte to a state or set of state(s) that are thermodynamically similar to globular proteins, one would expect to observe structural transitions when mechanical force is applied. Indeed, upon addition of TMAO, we observe structural transitions that we believe correspond to cooperative unfolding events (Fig. 5-6). Since we do not observe such transitions under these solution conditions when we pull on DNA handles only, we can conclude that this signal is coming specifically from the protein. Additionally, these types of cooperative transitions are very similar to what is observed for their globular protein counterparts [47, 106, 168], suggesting that these transitions correspond to tertiary structural unfolding transitions.

We observe such transitions for the A-NTD and C3-NTD, but not for the D1-NTD. This is a surprising result as the D1-NTD is predicted to have a substantial amount of structure from disorder prediction analysis [147]. This result alone is telling. It suggests that the minimal cooperative unit in the case of GR's NTD must contain a large portion of the molecule – i.e. one that at least contains the entire C3-NTD region. This may suggest an explanation for why certain disordered proteins do not behave cooperatively as expected when they are truncated unlike their globular protein counterparts [169].

The labeling chemistry and biochemistry developed in the previous chapter was able to reliably generate single-molecule tethers for our experiments. To further probe the energy landscape of GR's NTD, many data sets were collected for over 170 molecules of the A-NTD and C3-NTD as described in the following section.

Glucocorticoid Receptor's N-terminal domain is either misfolded or multi-state

Multiple force-ramp data sets were collected on molecules of A-NTD and C3-NTD. The results below will describe conclusions from the A-NTD, however, the same conclusions are drawn upon rigorous analysis of the C3-NTD and is not shown for brevity. Two main observations suggest that the protein is multi-state beyond the capability of being state assigned, or mis-folded.

The first main observation is that the majority of protein molecules do not show any structural transitions (Table 5-2 – rows 3 through 6). Specifically, approximately 67% or two-thirds of the molecules do not show any transitions upon addition of TMAO even when the concentration of TMAO is changed. These tethers are exhibiting static heterogeneity, i.e. individual tethers do not show transitions at all, or consistently refold and unfold in force-ramp experiments. Using the same metrics as describe previously to probe for protein, we are confident that the tethers that do not exhibit transitions contain protein and are not empty. It is also of note that the labeling and pulling experimental setup produces single-molecule tethers with protein for ~75% of bead pairs with all other proteins tested, and that the tethering efficiency under the conditions tested above is approximately the same. This suggests that we have not increased the odds of generating an "empty tether" significantly. Taken together, these results lead us to conclude that a significant portion of tethers contain protein, yet do not show structural transitions even when mechanically manipulated up to the overstretching transition.

The second main observation consistent with mis-folding or multi-state behavior occurs when we consider the force and change in extension of unfolding for every transition observed from force-ramp experiments (Fig. 5-7A). These results are in stark contrast to those obtained for T4 Lysozyme in the previous chapter. The results from Figure 5-7A are consistent with the two hypotheses discussed. These results have been observed before for other globular proteins that are multi-domain and multi-state under mechanical force (Christian Kaiser lab, unpublished data). Additionally, the number of conserved methionine start sites in the NTD suggests selective pressure to maintain these sites, which may or may not correspond to individual domains. In fact, even if a few of these isoforms corresponded to different functional domains, the results we obtain are exactly what we would expect. The other hypothesis that needs to be considered is that this may also be indicative of a misfolded protein.

To test the latter scenario, we hypothesized that proline isomerization may play a role. The full-length A-NTD has 23 prolines, therefore, mutagenesis is not practical. Additionally, bulk experiments carried out demonstrated that reaching the folded baseline in fluorsescence experiments takes on the order of minutes [54], consistent with the timescales where proline isomerization may be the rate limiting step [170]. We found waiting for proper folding to occur in the single-molecule experiment was not possible because of drift that occurs on the timescale of minutes.

Instead, we performed experiments in the presence of proline isomerse cyclophillin A (CypA). CypA has been previously demonstrated to assist in the proper folding of proteins by expediting the rate of folding when proline isomerization is be the rate limiting step [110, 170, 171]. The behavior of the single-molecules changed radically under these conditions. Specifically, molecules showed consistent refolding events and unfolding transitions were reproducible (Fig. 5-8A). These types of tethers are even amenable to constant-force experiments which clearly indicate cooperative refolding and unfolding events as we hop between low and

high forces (Fig. 5-8B). These results indicate that the transitions we previously observed may have come from mis-folded and possibly properly folded molecules. If this were the case, we would expect to observe similar unfolding transitions, and new transitions that were not observed before. Indeed this is the case (Fig. 5-9).

Two main conclusions can be drawn from Figure 5-9. The first is that we observe similar transitions to what has been seen before which is indicated by some of the overlap between the two data sets. The second, is that we are folding to different states which can be appreciated by the absence of certain transitions (bottom left corner), and the introduction of new transitions (bottom right corner). In fact, some of the transitions observed here are consistent with the full-length of the protein suggesting the entire molecule may fold cooperatively with its N-terminus on its C-terminus.

Unfortunately, the results contained in Figure 5-9 are still too complicated to be reliably state assigned. Additionally, the proportion of molecules showing transition is still low which suggests that certain molecules may still be misfolded (Table 5-2). As per the previous discussion, one would still expect to see complex multi-state behavior if the molecule was properly folded, consistent with what we observe. This conclusion is also valid for C3-NTD. The next logical step is to clearly reduce the level of complexity by pulling on different constructs and isoforms. This proposed mutagenic approach has been previously used to identify sub-domains in globular protein studies [47].

The most logical course of action for future work is to start with the structured DBD, and then add disordered segments to it such as the D1, D2, and D3 isoforms. This is a great starting point for deconvolution of the transitions because the DBD is known to be globular and will provide a well-defined structure [87], as well as a wealth of bulk-biochemical data [172], to map the transitions we observe on to. Additionally, since the DBD is positively coupled to the NTD

[147], we expect that this will stabilize the NTD and make the molecule more amenable to these experiments. It is our hope that the data presented in this chapter will be useful in informing results from future experiments.

5.5 Conclusions

The results presented here reveal new dimensions to coupling in disordered proteins, their thermodynamics of folding, and their experimental amenability to single-molecule techniques. Nature has clearly utilized the multiple isoforms of GR's NTD and tuned their stability via the delicate balance of states to modulate *in vivo* activity. It is also clear from the bulk-thermodynamic and structure based calculations that the NTD isoforms fold cooperatively to a state or set of state(s) that are thermodynamically indistinguishable to globular proteins. As such, one would expect to observe cooperative unfolding events under mechanical force at the single-molecule level. Indeed, this is what is observed when certain isoforms are folded in osmolyte and mechanically manipulated. As expected, the isoforms are multi-state and/or have multiple native states at the single-molecule level. Unfortunately, their complex behavior hinders reliable state assignment on the current constructs tested. Future directions include adding the globular DBD to the NTD and performing experiments on these constructs. It is likely that the current data will be informative once these experiments are conducted.

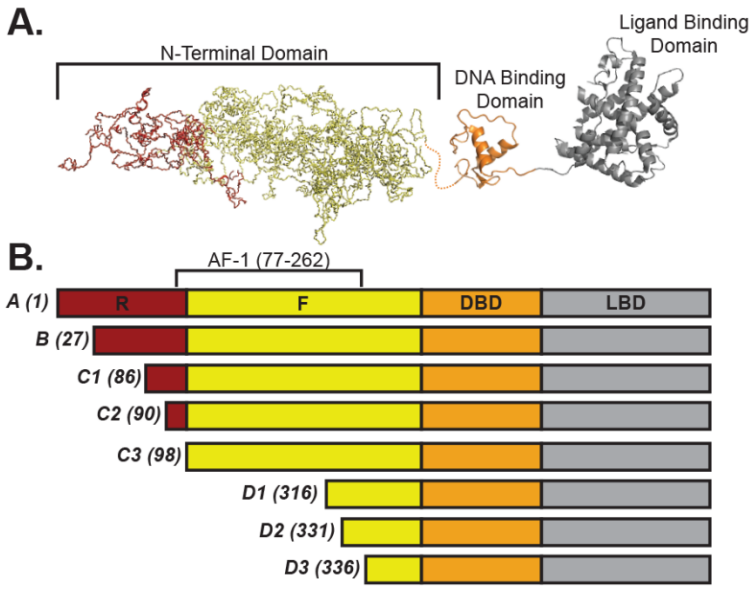

Figure 5-1. Glucocorticoid Receptor domain architecture and naturally occurring translational isoforms. **A**. Depicted from N-terminus to C-terminus is the domain architecture of GR. From left to right is the intrinsically disordered N-terminal domain (NTD), the DNA binding domain (DBD), and the Ligand Binding Domain (LBD). The NTD consist of two coupled domains: The regulatory (R) region and the Functional (F) region in red and yellow respectively. **B.** The eight naturally occurring translational isoforms of GR are depicted. Nomenclature is as defined before [154]. The activation function 1 (AF-1) domain is indicated on top of the isoforms as amino acids 77-262 [154]. Figure adopted from [173, 174].

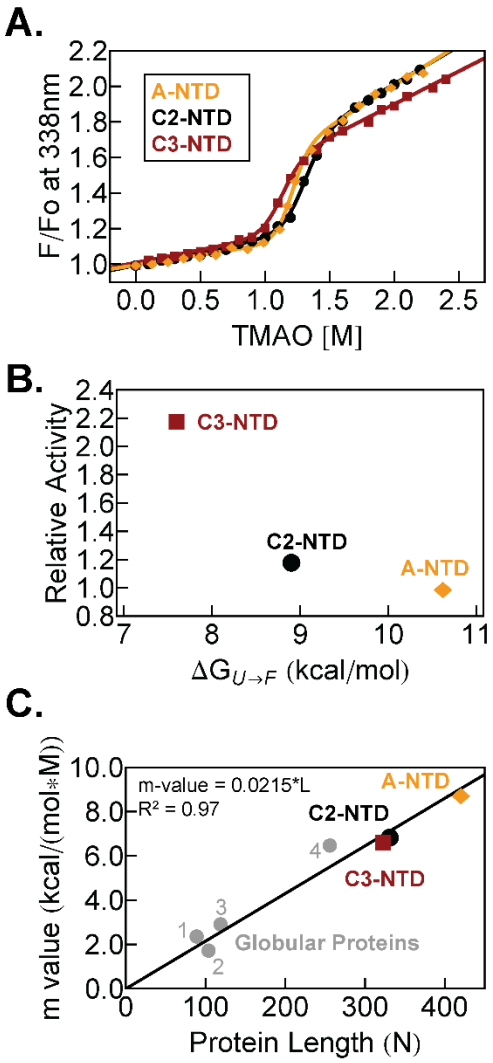

Figure 5-2. Glucocorticoid receptor N-Terminal Domain isoform stability contributes to *in vivo* activity and refolds to a globular conformation. **A**. Conformational transitions of GR NTD isoforms (A, C2, and C3) induced by TMAO. Transitions were monitored by tryptophan emission fluorescence intensity at 338nm with excitation at 295nm. Emission intensities were normalized to that at 0M TMAO. Detailed experimental conditions have been previously described [150]. **B.** Relative activities of GR isoforms [154] plotted against the experimentally measured stabilities ($\Delta G^{o}_{U\rightarrow F}$) from figure 5-2A. The plot indicates that stability correlates to relative activity

in a non-linear fashion. **C.** Experimentally measured m-values are plotted against the protein length (N) – i.e. number of amino acids. Shown are globular proteins (see [150] for more details) in numbered gray circles: (1) Barstar, (2) RCAM-T1, (3) P protein, and (4) Nank 1-7. The GR translational isoforms from Figure 5-2A fall on the same line as the globular proteins indicating that they are folding to a conformation that is thermodynamically indistinguishable from a globular fold [150]. Figure adopted from [54, 173].

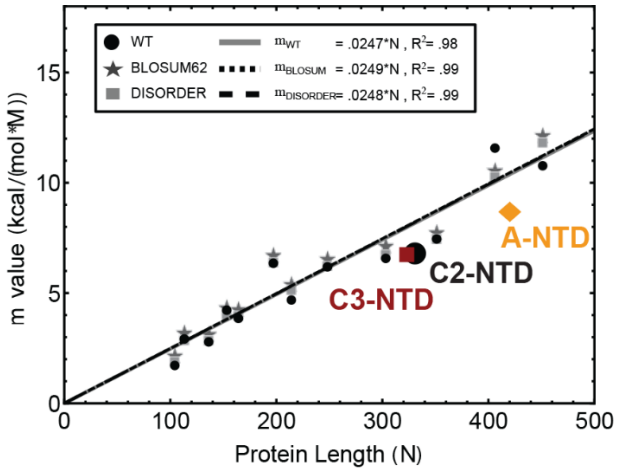

Figure 5-3. The refolded state(s) of disordered proteins are thermodynamically similar to globular proteins. Dependence of calculated m-values on protein length (N) and sequence composition. WT sequences are depicted in black circles, and can be fitted to a linear line (shown as gray line), i.e., m-value=0.0247*N, with a correlation coefficient of R^2=0.98. BLOSUM62 [160] mutated sequences are in dark gray stars, and can be fitted to a linear line (shown as black short dashed line), i.e., m-value=0.0249*N, with a correlation coefficient of R^2=0.99. DISORDER mutated sequences are in light gray boxes, and can be fitted to a linear line (shown as black long dashed line), i.e., m-value=0.0248*N, with a correlation coefficient of R^2=0.99. All three fits have similar dependence on N. The absolute changes in slope are only minutely dependent on variations in sequence composition because while ~25% of the m-value contribution comes from side chains contributions (with the remaining 75% coming from backbone contributions – Table 5-1), the relatively minor differences in amino acid usage between ID and globular proteins amount to only a small fraction of the computed m-value. For comparison, GR A-NTD, GR C2-NTD, and GR C3-NTD are depicted using the same color scheme as Figure 5-2. Figure adopted from [54].

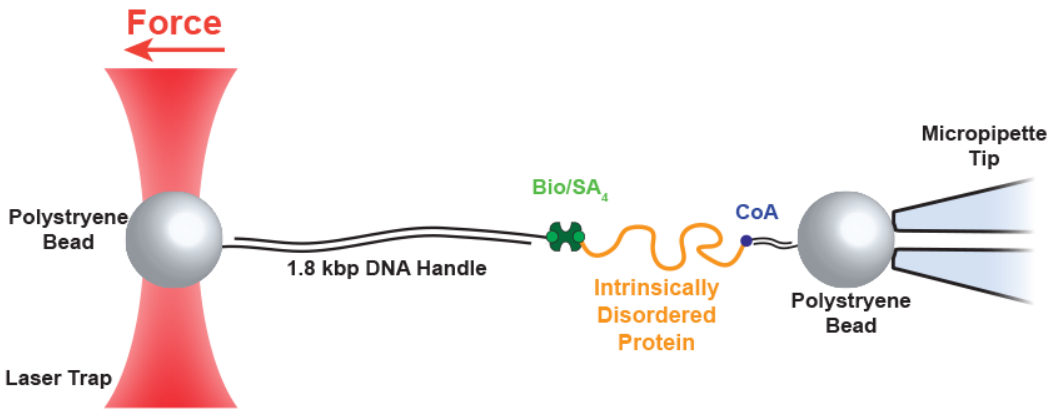

Figure 5-4. Single-molecule force spectroscopy experimental setup for intrinsically disordered proteins. In the experimental setup a polystyrene bead is held in the trap that exerts force on the single molecule. The bead is covalently linked to 1.8kbp of dsDNA which is attached to biotin (green circle) at the other end. There is a non-covalent biotin/streptatvidin interaction that then is linked to the IDP of choice. CoA (blue circle) is covalently linked to the disordered protein and another 50bp of dsDNA that is attached to the polystyrene bead which is held by suction.

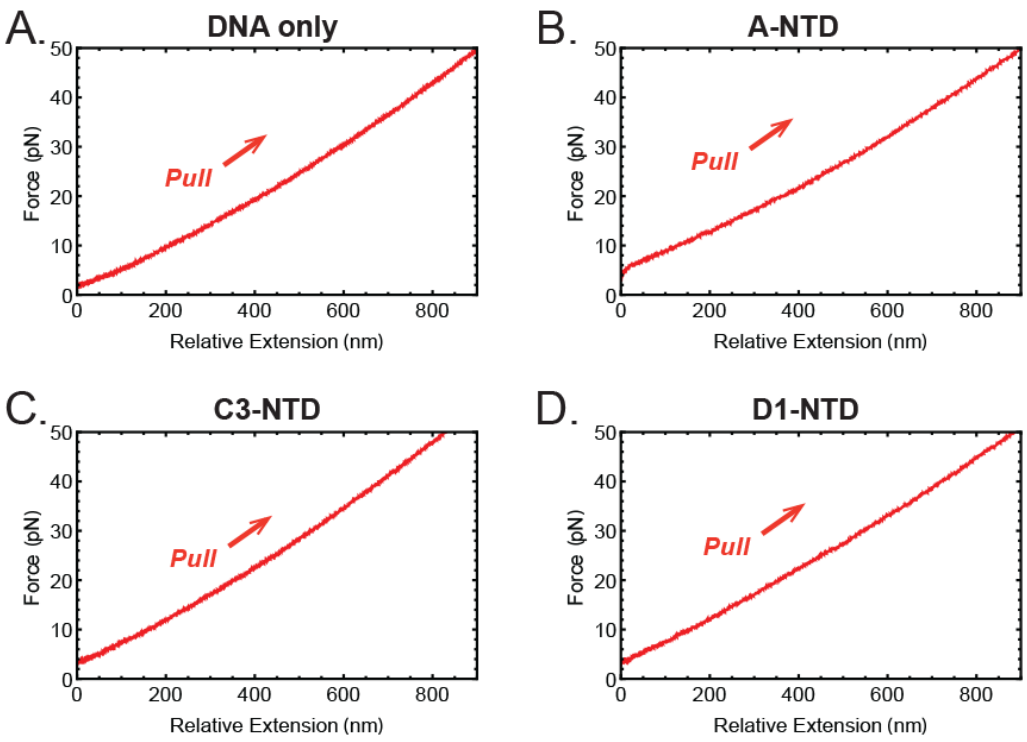

Figure 5-5. Intrinsically disordered proteins do not exhibit tertiary structural transitions under mechanical force. Representative force extension curves for a constant velocity experiment (100nm/s). Shown are data for DNA only (**A**), A-NTD (**B**), C3-NTD (**C**), and D1-NTD (**D**). What is clear from these data is that no significant structural transitions occur as expected. These results are reproducible from many different single-molecules.

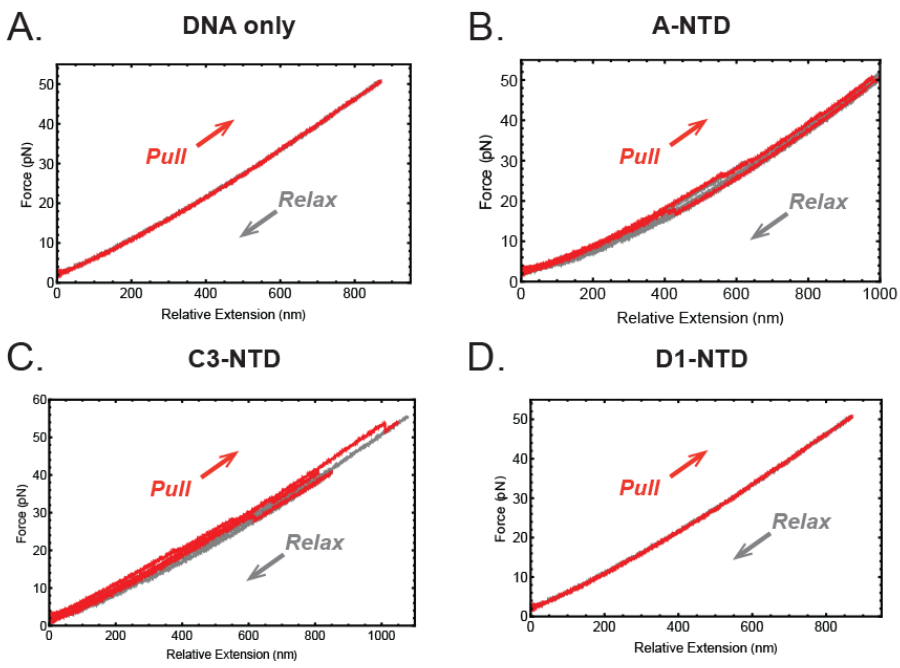

Figure 5-6. Certain isoforms of an intrinsically disordered protein exhibit mechanical unfolding transitions upon addition of osmolyte. Representative force extension curves for a constant velocity experiment (100nm/s). Shown are data for DNA only (**A**), A-NTD (**B**), C3-NTD (**C**), and D1-NTD (**D**). DNA and D1-NTD do not show any significant structural transitions even upon addition of 2M TMAO. However, the A-NTD and C3-NTD show clear structural transitions that result from the protein supporting the hypothesis that IDPs may behave like globular proteins under mechanical force.

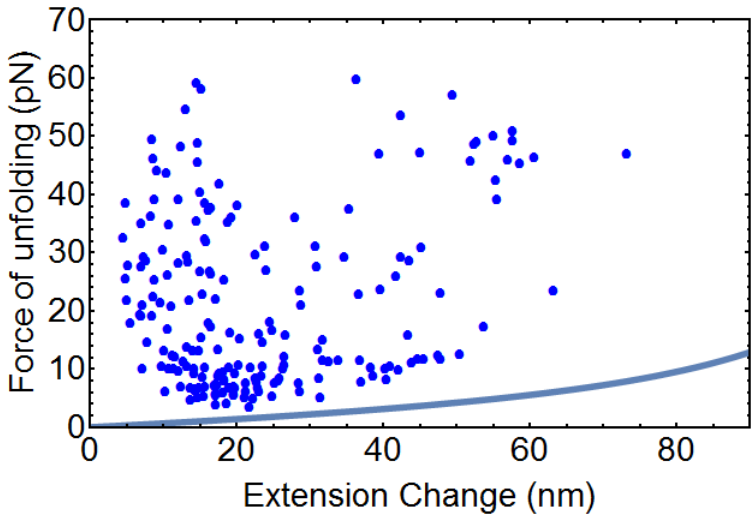

Figure 5-7. The TMAO refolded state(s) of A-NTD are heterogeneous when mechanically unfolded. Shown is a plot of every unfolding transition observed for the A-NTD in 2M TMAO. The force of unfolding (pN) is plotted against the extension change (nm). The blue line is a WLC-model for the full-length protein assuming the N- and C- termini are in physical contact (i.e. the maximum change we expect to observe). Since all transitions fall above the WLC-model line we believe that these all originate from single molecules (in addition to the other data discussed in the text). What is clear is that there is clear pattern in the data that would allow reliable state assignment. The distribution of points is not uniform, however, the data are too complicated to be state assigned.

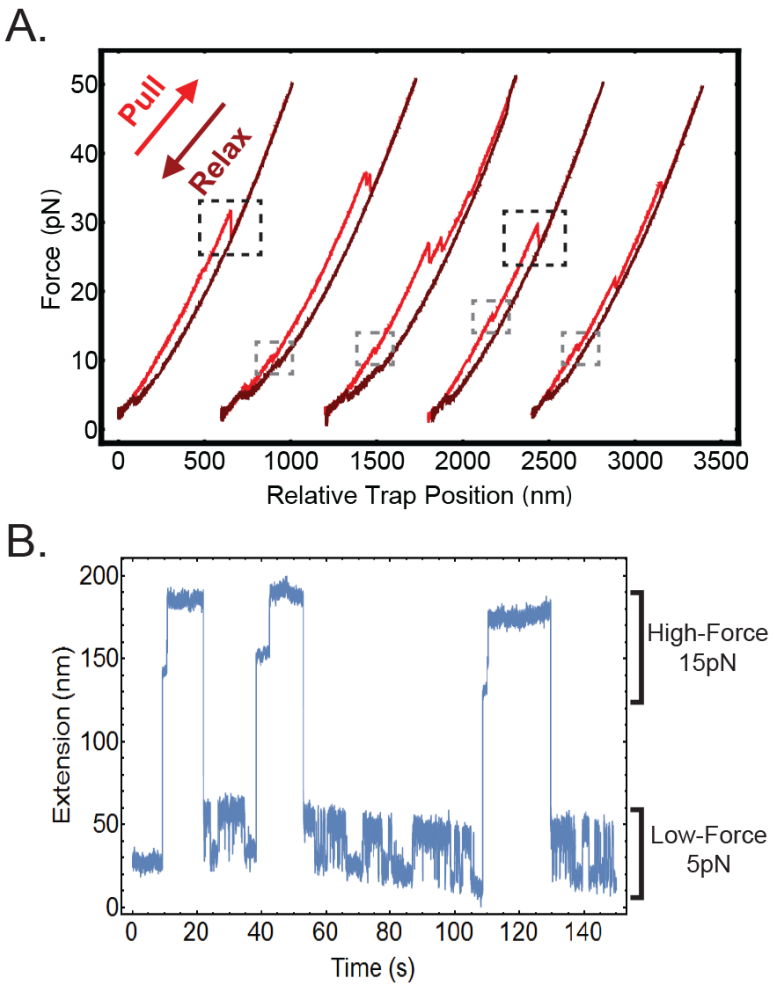

Figure 5-8. GR's NTD is misfolded upon addition of osmolyte. **A**. Shown are representative force-extension curves for five different molecules of A-NTD in the presence of 1μM proline isomerase. Highlighted in black and gray are transitions that reproducibly appear in the traces. Of note is that these transitions are of larger magnitude than those observed without proline isomerase present (see Fig. 5-6B). Of note is that there are certain transitions that are not found in all traces indicating multi-state behavior and or multi-domain behavior. Most importantly, all the molecule display refolding transitions in the low force regime during relaxing (dark red). **B.** Constant force data for a single-molecule of the A-NTD changing between 5pN (lower extension) and 15pN

(higher extension). Within 5pN, the molecule undergoes equilibrium "hopping" behavior between an ostensibly folded and unfolded state. At high force we observe multi-state unfolding transitions. These data indicate that addition of proline isomerase clearly leads to reproducible folding and unfolding transitions.

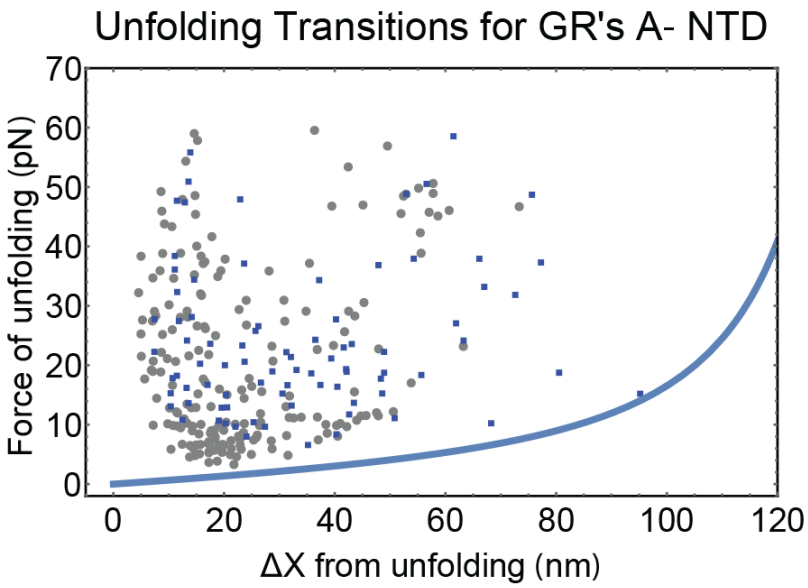

Figure 5-9. States of A-NTD is too heterogeneous to assign even in the presence of proline isomerase. Shown is a plot of every unfolding transition observed for the A-NTD in 3M TMAO and 1μM proline isomerase in blue versus the transitions in 2M TMAO from Figure 5-7. The blue line is a WLC-model for the full-length protein assuming the N- and C- termini are in physical contact (i.e. the maximum change we expect to observe). What is suggested is that certain transitions disappear (bottom left-corner) while new transitions appear (far right) that are near the full-length transition. These results suggest that the protein may have been misfolded without the addition of proline isomerase.

Table 5-1. Assessment of m-value contributions from protein backbone (BB) and side chains (SC)

Protein Name	PDB ID	Length[1]	m-value[2]	% BB[3]	% SC[3]	BLOSUM62[4]	DISORDER[4]
RNase T1	2BU4	104	1.78	70.3%	29.7%	1.79 ± 0.15	1.65 ± 0.10
P Protein	1A6F	113	2.99	75.3%	24.7%	2.67 ± 0.16	2.27 ± 0.10
SNase	1STN	136	2.85	71.8%	28.2%	2.70 ± 0.19	2.28 ± 0.13
Myoglobin	1MBN	153	4.29	77.8%	22.2%	3.73 ± 0.16	3.60 ± 0.13
T4 Lysozyme	2LZM	164	3.92	73.6%	26.4%	3.78 ± 0.19	3.37 ± 0.13
Carbonic Anhydrase	1CA2	197	6.42	73.2%	26.8%	5.64 ± 0.25	5.18 ± 0.19
Adenylate Kinase	1AKE	214	4.75	71.6%	28.4%	4.77 ± 0.22	4.23 ± 0.15
Triose Phospate Isomerase	1TIM	248	6.26	74.8%	25.2%	5.94 ± 0.23	5.19 ± 0.17
Porphobilinogen Deaminase	1PDA	303	6.64	71.8%	28.2%	6.89 ± 0.27	5.72 ± 0.19
Map Kinase	3QUD	351	7.52	72.2%	27.8%	6.29 ± 0.29	6.24 ± 0.21
Phosphoglycerate Dehydrogenase	2PA3	406	11.64	80.2%	19.8%	10.54 ± 0.28	9.62 ± 0.23
Seryl tRNA Synthetase	3QO8	451	10.84	71.4%	28.6%	11.57 ± 0.34	9.95 ± 0.23
Cytochrome C	1OCO	514	16.86	79.3%	20.7%	15.61 ± 0.35	15.28 ± 0.26
Amyloid Binding Protein	3DBH	526	13.40	74.2%	25.8%	12.82 ± 0.36	11.28 ± 0.25
			Average		**74.1%**	**25.9%**	
			Standard Deviation		**3.1%**	**3.1%**	

[1]Length was determined as the number of residues in the A-chain[2]m-values are all in kcal/(mol*M) and were calculated based on the methods described in the text.
[3]Percentages were calculated in absolute terms, i.e. $\%BB = |\Delta G_{tr,BB}|/(|\Delta G_{tr,BB}|+|\Delta G_{tr,SC}|)$ and $\%SC = |\Delta G_{tr,SC}|/(|\Delta G_{tr,BB}|+|\Delta G_{tr,SC}|)$
[4]Reported values are the averages and standard deviations of the 100,000 random mutated sequences generated.

Table 5-2. Statistics on single-molecules of A-NTD and their propensity for structural transitions under different conditions.

	Condition	0M TMAO	1.2M TMAO	2M TMAO	3M TMAO	3M TMAO + CypA	
# Molecules	Show transitions	0	5	20	9	8	
	Do not show transitions	34	9	40	12	33	**Total**
	Total	34	14	60	21	41	170
	% Transitions	0.00%	35.71%	33.33%	42.86%	19.51%	
	% No Transitions	100.00%	64.29%	66.67%	57.14%	80.49%	

References:

1. Yancey, P.H., et al., *Living with Water-Stress - Evolution of Osmolyte Systems.* Science, 1982. **217**(4566): p. 1214-1222.
2. Alberts, B., et al., *Molecular Biology of the Cell*. 2007: Garland Science.
3. Changeux, J.P., *Allosteric Mechanisms of Signal Transduction.* Science, 2005. **308**(5727): p. 1424-1428.
4. Changeux, J.P., *Allostery and the Monod-Wyman-Changeux Model After 50 Years.* Annual Review of Biophysics, 2012. **41**: p. 103-133.
5. Changeux, J.P., *The origins of allostery: from personal memories to material for the future.* Journal of Molecular Biology, 2013. **xxx**(xxx): p. xx-xx.
6. Cui, Q. and M. Karplus, *Allostery and cooperativity revisited.* Protein science : a publication of the Protein Society, 2008. **17**(8): p. 1295-307.
7. Fenton, A.W., *Allostery: an illustrated definition for the "second secret of life".* Trends in Biochemical Sciences, 2008. **33**(9): p. 420-425.
8. Hilser, V.J., J.O. Wrabl, and H. Motlagh, *Structural and energetic basis of allostery.* Annual Reviews of Biophysics, 2012. **41**: p. 585-609.
9. Nussinov, R., C.J. Tsai, and B. Ma, *The (still) underappreciated role of allostery in the cellular network.* Annual Reviews of Biophysics, 2013. **42**: p. xx-xx.
10. Nussinov, R. and C.J. Tsai, *Allostery in disease and drug discovery.* Cell, 2013. **153**: p. 293-305.
11. Monod, J. and F. Jacob, *Teleonomic mechanisms in cellular metabolism, growth, and differentiation.* Cold Spring Harbor Symposium on Quantitative Biology, 1961. **26**: p. 389-401.
12. Changeux, J.P., *The feedback control mechanisms of biosynthetic L-threonine deaminase by L-isoleucine.* Cold Spring Harbor Symposium on Quantitative Biology, 1961. **26**: p. 313-318.
13. Monod, J., J. Wyman, and J.P. Changeux, *On the nature of allosteric transitions: a plausible model.* Journal of Molecular Biology, 1965. **12**: p. 88-118.
14. M.F., P., *Stereochemistry of Cooperative Effects in Haemoglobin.* Nature, 1970. **228**: p. 726-734.
15. Perutz, M.F., *Stereochemistry of cooperative effects in haemoglobin.* Nature, 1970. **228**: p. 726-739.
16. Perutz, M.F., et al., *The stereochemical mechanism of the cooperative effects in hemoglobin revisited.* Annual Reviews of Biophysics and Biomolecular Structure, 1998. **27**: p. 1-34.
17. Perutz, M.F., et al., *Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-A resolution, obtained by x-ray analysis.* Nature, 1960. **185**: p. 416-422.
18. Daily, M.D. and J.J. Gray, *Local motions in a benchmark of allosteric proteins.* Proteins: Struct Funct Bioinf, 2007. **67**: p. 385-399.
19. Daily, M.D. and J.J. Gray, *Allosteric communication occurs via networks of tertiary and quaternary motions in proteins.* PLoS Comput Biol, 2009. **5**: p. e1000293.
20. Laskowski, R.A., F. Gerick, and J.M. Thornton, *The structural basis of allosteric regulation in proteins.* FEBS Letters, 2009. **583**: p. 1692-1698.
21. Changeux, J.P. and S.J. Edelstein, *Allosteric mechanisms of signal transduction.* Science, 2005. **308**: p. 1424-1428.
22. Dickerson, R.E., *X-ray studies of protein mechanisms.* Annual Reviews of Biophysical Chemistry, 1972. **41**: p. 815-842.
23. Gunasekaran, K., B. Ma, and R. Nussinov, *Is allostery an intrinsic property of all dynamic proteins?* Proteins: Struct Funct Bioinf, 2004. **57**(3): p. 433-443.

24. Tzeng, S.R. and C.G. Kalodimos, *Protein dynamics and allostery: an NMR view.* Curr Opin Struct Biol, 2011. **21**(1): p. 62-67.
25. Kern, D. and E.R. Zuiderweg, *The role of dynamics in allosteric regulation.* Curr Opin Struct Biol, 2003. **13**(6): p. 748-57.
26. Smock, R.G. and L.M. Gierasch, *Sending signals dynamically.* Science, 2009. **324**: p. 198-203.
27. Tsai, C.J., A. del Sol, and R. Nussinov, *Protein allostery, signal transmission and dynamics: a classification scheme of allosteric mechanisms.* Mol. BioSyst., 2009. **5**: p. 207-216.
28. Swain, J.F., et al., *Hsp70 chaperone ligands control domain association via an allosteric mechanism mediated by the interdomain linker.* Molecular Cell, 2007. **26**: p. 27-39.
29. Zuiderweg, E.R., et al., *Allostery in the Hsp70 chaperone proteins.* Topics in Current Chemistry, 2013. **328**: p. 99-153.
30. Petit, C.M., et al., *Hidden dynamic allostery in a PDZ domain.* Proc Natl Acad Sci U S A, 2009. **106**(43): p. 18249-18254.
31. Hilser, V.J. and E.B. Thompson, *Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins.* Proc Natl Acad Sci U S A, 2007. **104**(20): p. 8311-5.
32. Motlagh, H.N., et al., *The ensemble nature of allostery.* Nature, 2014. **508**: p. 331-339.
33. Monod, J., Wyman, J., Changeux, J.P., *On the Nature of Allosteric Transitions: A Plausible Model.* Journal of Molecular Biology, 1965. **12**: p. 88-118.
34. Koshland, D.E., Nemethy, G., and Filmer. D., *Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits.* Biochemistry, 1966. **5**(1): p. 365-385.
35. Hilser, V.J. and E.B. Thompson, *Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins.* Proceedings of the National Academy of Sciences, 2007. **104**(20): p. 8311-8315.
36. Koshland, D.E., Jr., G. Nemethy, and D. Filmer, *Comparison of experimental binding data and theoretical models in proteins containing subunits.* Biochemistry, 1966. **5**(1): p. 365-85.
37. Monod, J., J. Wyman, and J.P. Changeux, *On the Nature of Allosteric Transitions: A Plausible Model.* J Mol Biol, 1965. **12**: p. 88-118.
38. Branden, C. and J. Tooze, *Introduction to Protein Structure*. 2 ed. 1999, New York: Garland Science.
39. Wyman, J. and S. Gill, *Binding and linkage: functional chemistry of biological macromolecules.* 1990.
40. Wyman, J., Gill, S. J., *Binding and Linkage*. 1 ed. 1990: University Science Books.
41. Hilser, V.J., and Thompson EB, *Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins*. Proceedings of the National Academy of Sciences USA, 2007. **104**(20): p. 8311-8315.
42. Gielen, S.C., et al., *Analysis of Estrogen Agonism and Antagonism of Tamoxifen, Raloxifene, and ICI182780 in Endometrial Cancer Cells: A Putative Role for the Epidermal Growth Factor Receptor Ligand Amphiregulin.* Society for Gynecologic Investigation, 2005. **12**: p. 55-66.
43. Ferreon, A.C.M., et al., *Modulation of allostery by protein intrinsic disorder.* Nature, 2013. **498**: p. 390-394.
44. Serdyuk, I.N., N.R. Zacai, and J. Zaccai, *Methods in Molecular Biophysics - Structure, Dynamics, Function*. 2007: Cambridge University Press.
45. Uversky, V.N., J. Li, and A.L. Fink, *Trimethylamine-N-oxide-induced folding of alpha-synuclein.* FEBS Lett, 2001. **509**(1): p. 31-5.
46. Shank, E.A., et al., *The folding cooperativity of a protein is controlled by its chain topology.* Nature, 2010. **465**(7298): p. 637-U134.

47. Gao, Y., et al., *Single reconstituted neuronal SNARE complexes zipper in three distinct stages.* Science, 2012. **337**(6100): p. 1340-3.
48. Kaiser, C.M., et al., *The Ribosome Modulates Nascent Protein Folding.* Science, 2011. **334**(6063): p. 1723-1727.
49. Auton, M. and D.W. Bolen, *Predicting the energetics of osmolyte-induced protein folding/unfolding.* Proc Natl Acad Sci U S A, 2005. **102**(42): p. 15065-8.
50. Auton, M. and D.W. Bolen, *Application of the Transfer Model to Understanding How Naturally Occuring Osmolytes Affect Protein Stability.* Methods in Enzymology, 2007. **428**: p. 397-418.
51. Auton, M., et al., *Osmolyte effects on protein stability and solubility: a balancing act between backbone and side-chains.* Biophysical chemistry, 2011. **159**(1): p. 90-9.
52. Liu, J., Perumal NB, Oldfield CJ, Su EW, Uversky VU, Dunker AK, *Intrinsic Disorder in Transcription Factors.* Biochemistry, 2006. **45**: p. 6873-6888.
53. Sevcsik, E., et al., *Allostery in a disordered protein: oxidative modifications to alpha-synuclein act distally to regulate membrane binding.* Journal of the American Chemical Society, 2011. **133**: p. 7152-7158.
54. Li, J., et al., *Thermodynamic Dissection of the Intrinsically Disordered n-Terminal Domain of Human Glucocorticoid Receptor.* Journal of Biological Chemistry, 2012: p. (In Press).
55. Motlagh, H. and V.J. Hilser, *Agonsim/antagonism switching in allosteric ensembles.* Proceedings of the National Academy of Sciences, U.S.A., 2012. **2012**: p. 4134-4139.
56. Courter, J.R., et al., *Structure-based design, synthesis and validation of CD4-mimetic small molecule inhibitors of HIV-1 entry: conversion of a viral entry agonist to an antagonist.* Acc Chem Res, 2014. **47**(4): p. 1228-37.
57. Popovych, N., et al., *Dynamically driven protein allostery.* Nat Struct Mol Biol, 2006. **13**: p. 831-838.
58. Tzeng, S.R. and C.G. Kalodimos, *Dynamic activation of an allosteric regulatory protein.* Nature, 2009. **462**(7271): p. 368-72.
59. Tzeng, S.R. and C.G. Kalodimos, *Protein activity regulation by conformational entropy.* Nature, 2012. **488**: p. 236-240.
60. Freiburger, L.A., et al., *Competing allosteric mechanisms modulate substrate binding in a dimeric enzyme.* Nature Structural & Molecular Biology, 2011. **18**: p. 288-294.
61. Reichheld, S.E., Z. Yu, and A.R. Davidson, *The induction of folding cooperativity by ligand binding drives the allosteric response of tetracycline repressor.* Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(52): p. 22263-8.
62. Wright, P.E., Dyson JH, *Intrinsically Unstructured Proteins: Re-assessing the Protein Structure-function Paradigm.* Journal of Molecular Biology, 1999. **293**: p. 321-331.
63. Uversky, V.U., Oldfiend CJ, Dunker AK, *Showing your ID: intrinsic disorder as an Id for recognition, regulation and cell signaling.* Journal of Molecular Recognition, 2005. **18**: p. 343-384.
64. Dunker, A.K., et al, *Intrinsically disordered protein.* Journal of Molecular Graphics and Modelling, 2001. **19**: p. 26-59.
65. Babu, M.M., Lee R, de Groot NS, and Gsponer J, *Intrinsically disordered proteins: regulation and disease.* Current Opinion in Structural Biology, 2011. **21**: p. 432-440.
66. Bai, F., et al., *Conformational spread as a mechanism for cooperativity in the bacterial flagellar switch.* Science, 2010. **327**(5966): p. 685-9.
67. Clarkson, M.W., et al., *Dynamic coupling and allosteric behavior in a nonallosteric protein.* Biochemistry, 2006. **45**(25): p. 7693-9.
68. Cooper, A. and D.T. Dryden, *Allostery without conformational change. A plausible model.* European biophysics journal : EBJ, 1984. **11**(2): p. 103-9.

69. Laine, O., et al., *Allosteric signaling in the biotin repressor occurs via local folding coupled to global dampening of protein dynamics.* Journal of molecular biology, 2008. **381**(1): p. 89-101.
70. Pan, H., J.C. Lee, and V.J. Hilser, *Binding sites in Escherichia coli dihydrofolate reductase communicate by modulating the conformational ensemble.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(22): p. 12020-5.
71. Smock, R.G. and L.M. Gierasch, *Sending signals dynamically.* Science, 2009. **324**(5924): p. 198-203.
72. Stratton, M.M., Loh SN, *On the mechanism of protein fold-switching by a molecular sensor.* Proteins: Structure Function Bioinformatics, 2010. **78**: p. 3260-3269.
73. Garza, A.S., et al., *Binding-Folding Induced Regulation of AF1 Transactivation Domain of the Glucocorticoid Receptor by a Cofactor That Binds to Its DNA Binding Domain.* PLoS ONE, 2011. **6**(10): p. e25875.
74. Kenakin, T.P., *'7TM receptor allostery: putting numbers to shapeshifting proteins.* Trends in pharmacological sciences, 2009. **30**(9): p. 460-9.
75. Kenakin, T. and L.J. Miller, *Seven transmembrane receptors as shapeshifting proteins: the impact of allosteric modulation and functional selectivity on new drug discovery.* Pharmacological reviews, 2010. **62**(2): p. 265-304.
76. Latchman, D.S., *Transcription Factors: An Overview.* International Journal of Biochemistry and Cell Biology, 1997. **29**: p. 1305-1312.
77. Hol, T., Cox MB, Bryant HU, Draper MW, *Selective Estrogen Receptor Modulators and Postmenopausal Women's Health.* Journal of Women's Health, 1997. **6**: p. 523-531.
78. Katzenellenbogen, B.S., et al., *Antiestrogens: Mechanisms of action and resistance in breast cancer.* Breast Cancer Research and Treatment, 1997. **44**: p. 23-38.
79. Lu, N.Z., Cidlowski JA, *Translational Regulatory Mechanisms Generate N-Terminal Glucocorticoid Receptor Isoforms with Unique Transcriptional Target Genes.* Molecular Cell, 2005. **18**: p. 331-342.
80. Koshland, D.E., G. Nemethy, and D. Filmer, *Comparison of experimental binding data and theoretical models in proteins containing subunits.* Biochemistry, 1966. **5**: p. 365-385.
81. Cesareni, G., et al., eds. *Modular Protein Domains*. 2005, Wiley-VCH: Weinheim, FRG.
82. Wyman J, G.S., *Binding and Linkage*. 10 ed. 1990: University Science Books.
83. Schrank, T.P., D.W. Bolen, and V.J. Hilser, *Rational modulation of conformational fluctuations in adenylate kinase reveals a local unfolding mechanism for allostery and functional adaptation in proteins.* Proc Natl Acad Sci U S A, 2009. **106**(40): p. 16984-9.
84. Schrank, T.P., et al., *Strategies for the thermodynamic characterization of linked binding / local folding reactions within the native state: application to the LID domain of adenylate kinase from Escherichia coli.* Methods in Enzymology, 2011. **492**: p. 253-282.
85. Gekko, K., et al., *A linear correlation between the energetics of allosteric communication and protein flexibility in the Escherichia coli cyclic AMP receptor protein revealed by mutation-induced changes in compressibility and amide hydrogen-deuterium exchange.* Biochemistry, 2004. **43**(13): p. 3844-52.
86. Luque, I., S.A. Leavitt, and E. Freire, *The linkage between protein folding and functional cooperativity: two sides of the same coin?* Annual review of biophysics and biomolecular structure, 2002. **31**: p. 235-56.
87. Meijsing, S.H., et al., *DNA binding site sequence directs glucocorticoid receptor structure and activity.* Science, 2009. **324**(5925): p. 407-10.
88. Ahmad, A., Kumar R, *Steroid hormone receptors in cancer development: A target for cancer therapeutics.* Cancer Letters, 2011. **300**: p. 1-9.

89. Ashkin, A., et al., *Observation of a Single-Beam Gradient Force Optical Trap for Dielectric Particles.* Optics Letters, 1986. **11**(5): p. 288-290.
90. Ashkin, A., *Acceleration and Trapping of Particles by Radiation Pressure.* Physical Review Letters, 1970. **24**(4): p. 156-&.
91. Bustamante, C., et al., *Mechanical processes in biochemistry.* Annual Review of Biochemistry, 2004. **73**: p. 705-748.
92. Bustamante, C., et al., *Entropic elasticity of lambda-phage DNA.* Science, 1994. **265**(5178): p. 1599-600.
93. Tans, S., et al., *Single molecule studies of the DNA packaging motor of bacteriophage Phi 29.* Biophysical Journal, 2001. **80**(1): p. 200a-201a.
94. Smith, D.E., et al., *The bacteriophage phi 29 portal motor can package DNA against a large internal force.* Nature, 2001. **413**(6857): p. 748-752.
95. Smith, S.B., Y. Cui, and C. Bustamante, *Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules.* Science, 1996. **271**(5250): p. 795-9.
96. Liphardt, J., et al., *Reversible unfolding of single RNA molecules by mechanical force.* Science, 2001. **292**(5517): p. 733-737.
97. Cecconi, C., et al., *Direct observation of the three-state folding of a single protein molecule.* Science, 2005. **309**(5743): p. 2057-2060.
98. Smith, S.B., Y. Cui, and C. Bustamante, *Optical-trap force transducer that operates by direct measurement of light momentum.* Methods Enzymol, 2003. **361**: p. 134-62.
99. Wang, G.M., et al., *Experimental demonstration of violations of the second law of thermodynamics for small systems and short time scales.* Physical Review Letters, 2002. **89**(5).
100. Evans, D.J., E.G.D. Cohen, and G.P. Morriss, *Probability of 2nd Law Violations in Shearing Steady-States.* Physical Review Letters, 1993. **71**(15): p. 2401-2404.
101. Levine, I.N., *Physical chemistry*. 6th ed. 2009, Boston: McGraw-Hill. xviii, 989 p.
102. Fuller, D.N., et al., *A general method for manipulating DNA sequences from any organism with optical tweezers.* Nucleic Acids Res, 2006. **34**(2): p. e15.
103. Guinn, E.J., B. Jagannathan, and S. Marqusee, *Single-molecule chemo-mechanical unfolding reveals multiple transition state barriers in a small single-domain protein.* Nat Commun, 2015. **6**: p. 6861.
104. Jagannathan, B. and S. Marqusee, *Protein folding and unfolding under force.* Biopolymers, 2013. **99**(11): p. 860-869.
105. Jagannathan, B., et al., *Direct observation of a force-induced switch in the anisotropic mechanical unfolding pathway of a protein.* Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(44): p. 17820-17825.
106. Maillard, R.A., et al., *ClpX(P) generates mechanical force to unfold and translocate its protein substrates.* Cell, 2011. **145**(3): p. 459-69.
107. Gebhardt, J.C., T. Bornschlogl, and M. Rief, *Full distance-resolved folding energy landscape of one single protein molecule.* Proc Natl Acad Sci U S A, 2010. **107**(5): p. 2013-8.
108. Stigler, J., et al., *The complex folding network of single calmodulin molecules.* Science, 2011. **334**(6055): p. 512-6.
109. Street, T.O., D.W. Bolen, and G.D. Rose, *A molecular mechanism for osmolyte-induced protein stability.* Proc Natl Acad Sci U S A, 2006. **103**(38): p. 13997-4002.
110. Russo, A.T., J. Rosgen, and D.W. Bolen, *Osmolyte effects on kinetics of FKBP12 C22A folding coupled with prolyl isomerization.* Journal of molecular biology, 2003. **330**(4): p. 851-66.

111. Pradeep, L. and J.B. Udgaonkar, *Osmolytes induce structure in an early intermediate on the folding pathway of barstar.* Journal of Biological Chemistry, 2004. **279**(39): p. 40303-13.
112. Mukaiyama, A., et al., *Osmolyte effect on the stability and folding of a hyperthermophilic protein.* Proteins, 2008. **71**(1): p. 110-8.
113. Lin, S.L., A. Zarrine-Afsar, and A.R. Davidson, *The osmolyte trimethylamine-N-oxide stabilizes the Fyn SH3 domain without altering the structure of its folding transition state.* Protein Sci, 2009. **18**(3): p. 526-36.
114. Garcia-Manyes, S., L. Dougan, and J.M. Fernandez, *Osmolyte-induced separation of the mechanical folding phases of ubiquitin.* Proc Natl Acad Sci U S A, 2009. **106**(26): p. 10540-5.
115. Aioanei, D., et al., *Single-molecule-level evidence for the osmophobic effect.* Angew Chem Int Ed Engl, 2011. **50**(19): p. 4394-7.
116. Aioanei, D., et al., *Observing the osmophobic effect in action at the single molecule level.* Proteins, 2011. **79**(7): p. 2214-23.
117. Aioanei, D., et al., *Worm-like Ising model for protein mechanical unfolding under the effect of osmolytes.* Biophys J, 2012. **102**(2): p. 342-50.
118. Matsumura, M. and B.W. Matthews, *Control of enzyme activity by an engineered disulfide bond.* Science, 1989. **243**(4892): p. 792-4.
119. Parker, M.J. and S. Marqusee, *The Cooperativity of Burst Phase Reactions Explored.* Journal of Molecular Biology, 1999. **293**: p. 1195-1210.
120. Cellitti, J., R. Bernstein, and S. Marqusee, *Exploring subdomain cooperativity in T4 lysozyme II: uncovering the C-terminal subdomain as a hidden intermediate in the kinetic folding pathway.* Protein science : a publication of the Protein Society, 2007. **16**(5): p. 852-62.
121. Beckett, D., E. Kovaleva, and P.J. Schatz, *A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation.* Protein science : a publication of the Protein Society, 1999. **8**(4): p. 921-9.
122. Yin, J., et al., *Site-specific protein labeling by Sfp phosphopantetheinyl transferase.* Nat Protoc, 2006. **1**(1): p. 280-5.
123. Yin, J., et al., *Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase.* Proc Natl Acad Sci U S A, 2005. **102**(44): p. 15815-20.
124. Smith, S.B., L. Finzi, and C. Bustamante, *Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads.* Science, 1992. **258**(5085): p. 1122-6.
125. Lee, B. and F.M. Richards, *The interpretation of protein structures: estimation of static accessibility.* Journal of molecular biology, 1971. **55**(3): p. 379-400.
126. Phillips, J.C., et al., *Scalable molecular dynamics with NAMD.* J Comput Chem, 2005. **26**(16): p. 1781-802.
127. Best, R.B., et al., *Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles.* J Chem Theory Comput, 2012. **8**(9): p. 3257-3273.
128. Hilser, V.J. and E. Freire, *Structure-based calculation of the equilibrium folding pathway of proteins. Correlation with hydrogen exchange protection factors.* Journal of molecular biology, 1996. **262**(5): p. 756-72.
129. Dudko, O.K., G. Hummer, and A. Szabo, *Theory, analysis, and interpretation of single-molecule force spectroscopy experiments.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(41): p. 15755-15760.
130. Ando, N., et al., *Structural and thermodynamic characterization of T4 lysozyme mutants and the contribution of internal cavities to pressure denaturation.* Biochemistry, 2008. **47**(42): p. 11097-109.

131. Hilser, V.J., B.D. Townsend, and E. Freire, *Structure-based statistical thermodynamic analysis of T4 lysozyme mutants: Structural mapping of cooperative interactions.* Biophysical chemistry, 1997. **64**(1-3): p. 69-79.
132. Bell, G.I., *Models for the specific adhesion of cells to cells.* Science, 1978. **200**(4342): p. 618-27.
133. Chodera, J.D., et al., *Bayesian hidden Markov model analysis of single-molecule force spectroscopy: Characterizing kinetics under measurement uncertainty.* arXiv, 2011. **1108.1430**.
134. Cheng, C.L., M.Z. Zhang, and G.J. Zhao, *Mechanical stability and thermal conductivity of beta-barrel in green fluorescent protein by steered molecular dynamics.* Rsc Advances, 2014. **4**(13): p. 6513-6516.
135. Zheng, W., *All-atom and coarse-grained simulations of the forced unfolding pathways of the SNARE complex.* Proteins, 2014. **82**(7): p. 1376-86.
136. Zhang, Y. and J. Lou, *The Ca(2+) influence on calmodulin unfolding pathway: a steered molecular dynamics simulation study.* PLoS ONE, 2012. **7**(11): p. e49013.
137. Matthews, B.W. and S.J. Remington, *The three dimensional structure of the lysozyme from bacteriophage T4.* Proc Natl Acad Sci U S A, 1974. **71**(10): p. 4178-82.
138. Llinas, M., et al., *The energetics of T4 lysozyme reveal a hierarchy of conformations.* Nature structural biology, 1999. **6**(11): p. 1072-8.
139. Qu, Y. and D.W. Bolen, *Hydrogen exchange kinetics of RNase A and the urea:TMAO paradigm.* Biochemistry, 2003. **42**(19): p. 5837-49.
140. Tantos, A., K.H. Han, and P. Tompa, *Intrinsic disorder in cell signaling and gene transcription.* Mol Cell Endocrinol, 2012. **348**(2): p. 457-465.
141. Liu, J., et al., *Intrinsic disorder in transcription factors.* Biochemistry, 2006. **45**(22): p. 6873-88.
142. Tompa, P., *Intrinsically unstructured proteins.* Trends Biochem Sci, 2002. **27**(10): p. 527-33.
143. Uversky, V.N., *Intrinsically disordered proteins from A to Z.* Int J Biochem Cell Biol, 2011. **43**(8): p. 1090-103.
144. Dyson, H.J. and P.E. Wright, *Intrinsically unstructured proteins and their functions.* Nat Rev Mol Cell Biol, 2005. **6**(3): p. 197-208.
145. Fink, A.L., *Natively unfolded proteins.* Curr Opin Struct Biol, 2005. **15**(1): p. 35-41.
146. Dyson, H.J. and P.E. Wright, *Coupling of folding and binding for unstructured proteins.* Curr Opin Struct Biol, 2002. **12**(1): p. 54-60.
147. Li, J., *PhD Thesis Dissertation - Intrinsic Disorder and Allostery in Glucocorticoid Receptor*, in *T.C. Jenkins Department of Biophysics*. 2014, The Johns Hopkins University.
148. Babu, M.M., et al., *Intrinsically disordered proteins: regulation and disease.* Curr Opin Struct Biol. **21**(3): p. 432-40.
149. Midic, U., et al., *Unfoldomics of human genetic diseases: illustrative examples of ordered and intrinsically disordered members of the human diseasome.* Protein Pept Lett, 2009. **16**(12): p. 1533-47.
150. Li, J., et al., *Thermodynamic Dissection of the Intrinsically Disordered N-terminal Domain of Human Glucocorticoid Receptor.* The Journal of biological chemistry, 2012. **287**(32): p. 26777-87.
151. Garcia-Pino, A., et al., *Allostery and intrinsic disorder mediate transcription regulation by conditional cooperativity.* Cell, 2010. **142**: p. 101-111.
152. Lavery, D.N. and I.J. McEwan, *Structure and function of steroid receptor AF1 transactivation domains: induction of active conformations.* Biochem J, 2005. **391**(Pt 3): p. 449-64.

153. Lavery, D.N. and I.J.M.C. Ewan, *Structure and function of steroid receptor AF1 transactivation domains: induction of active conformations.* Biochemical Journal, 2005. **391**(449-464).
154. Lu, N.Z. and J.A. Cidlowski, *Translational regulatory mechanisms generate N-terminal glucocorticoid receptor isoforms with unique transcriptional target genes.* Molecular cell, 2005. **18**(3): p. 331-42.
155. Kumar, R. and E.B. Thompson, *Influence of flanking sequences on signaling between the activation function AF1 and DNA-binding domain of the glucocorticoid receptor.* Archives of Biochemistry and Biophysics, 2010. **496**: p. 140-145.
156. Kumar, R., et al., *Interdomain Signaling in a Two-domain Fragment of the Human Glucocorticoid Receptor.* J Biol Chem, 1999. **274**: p. 24737-24741.
157. Tanford, C., *Isothermal unfolding of globular proteins in aqueous urea solutions.* J. Am. Chem. Soc., 1964. **86**: p. 2050-2059.
158. Auton, M. and D.W. Bolen, *Predicting the energetics of osmolyte-induced protein folding/unfolding.* Proc Natl Acad Sci U S A, 2005. **102**(42): p. 15065-8.
159. Auton, M. and D.W. Bolen, *Application of the transfer model to understand how naturally occurring osmolytes affect protein stability.* Methods Enzymol, 2007. **428**: p. 397-418.
160. Henikoff, S. and J.G. Henikoff, *Amino acid substitution matrices from protein blocks.* Proceedings of the National Academy of Sciences of the United States of America, 1992. **89**(22): p. 10915-9.
161. Midic, U., A.K. Dunker, and Z. Obradovic, *Protein sequence alignment and structural disorder: a substitution matrix for an extended alphabet.* StReBio, 2009. **09**: p. 27-31.
162. Gibson, D.G., *Enzymatic assembly of overlapping DNA fragments.* Methods Enzymol, 2011. **498**: p. 349-61.
163. Gibson, D.G., et al., *Enzymatic assembly of DNA molecules up to several hundred kilobases.* Nat Methods, 2009. **6**(5): p. 343-5.
164. Tantos, A., K.H. Han, and P. Tompa, *Intrinsic disorder in cell signaling and gene transcription.* Molecular nd Cellular Endocrinology, 2012. **348**(2): p. 457-465.
165. Kumar, R., et al., *PROTEIN STRUCTURE AND FOLDING : The Conformation of the Glucocorticoid Receptor AF1 / tau1 Domain Induced by Osmolyte Binds Co-regulatory Proteins The Conformation of the Glucocorticoid Receptor AF1 / tau1 Domain Induced by Osmolyte Binds Co-regulatory Proteins.* The Journal of biological chemistry, 2001. **276**(21): p. 18146-18152.
166. Street, T.O., N. Courtemanche, and D. Barrick, *Protein Folding and Stability Using Denaturants.* Methods in Cell Biology, 2008. **84**: p. 295-325.
167. Uversky, V.N., *Natively unfolded proteins: a point where biology waits for physics.* Protein science : a publication of the Protein Society, 2002. **11**(4): p. 739-56.
168. Kaiser, C., et al., *The Ribosome Modulates Nascent Protein Folding.* Biophysical Journal, 2012. **102**(3): p. 68a-68a.
169. Sigler, P.B., *Transcriptional activation. Acid blobs and negative noodles.* Nature, 1988. **333**(6170): p. 210-2.
170. Bradley, C.M. and D. Barrick, *Effect of multiple prolyl isomerization reactions on the stability and folding kinetics of the notch ankyrin domain: experiment and theory.* J Mol Biol, 2005. **352**(2): p. 253-65.
171. Fraser, J.S., et al., *Hidden alternative structures of proline isomerase essential for catalysis.* Nature, 2009. **462**: p. 669-673.
172. Hudson, W.H., C. Youn, and E.A. Ortlund, *The structural basis of direct glucocorticoid-mediated transrepression.* Nat Struct Mol Biol, 2013. **20**(1): p. 53-8.
173. Motlagh, H.N., et al., *Disordered allostery: lessons from glucocorticoid recptor.* Biophysical Reviews, 2015. **7**(2): p. 257-65.

174. Motlagh, H.N., et al., *Interplay between allostery and intrinsic disorder in an ensemble.* Biochem Soc Trans, 2012. **40**(5): p. 975.

Vita

Hesam N. Motlagh was born in San Francisco, California on June 6th, 1988. He grew up in the most beautiful city in all the land: Cleveland, Ohio. He attended Miami University in Ohio where he spent his summers doing research in the lab of Dr. Chris Makaroff – department chair of chemistry and biochemistry. There he was able to learn about computational modeling of proteins and perform experiments to test his hypotheses. Falling in love with the science, he knew he wanted to pursue a doctoral degree in an interdisciplinary field. Graduating in 2010, he received a Bachelor of Science degree in Biochemistry, Bachelor of Science degree in Mathematics & Statistics, and minored in Actuarial Science and Molecular Biology. He immediately entered graduate school, wishing to integrate his pure math background with biology, and joined the Program in Molecular Biophysics at The Johns Hopkins University in Baltimore, Maryland. Pursuing research in both computational theory and experimental work, he chose Vincent J. Hilser as his mentor for his thesis work. At the time of this writing, Hesam is 27 years young, still single because he is married to his work, and has been working part time at a hedge-fund performing quantitative research. Upon completion of this thesis, he will continue being a part-time hedge-fund analyst, work at an investment management firm doing equity research, and also be a fellow in the Johns Hopkins Institute for Applied Economics, Global Health, and the Study of Business Enterprise. He will be teaching his first class next semester entitled, "Applied Economics and Finance," where he will teach students how to perform equity research under the mentorship of Professor Steve Hanke. He hopes to make enough money to endow the T. C. Jenkins Department of Biophysics chair position as the, "Hesam Motlagh Chair of Biophysics."